S/N 09/903,412 PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant	: Shohei Koide	Art Unit	: 1639
Serial No.	: 09/903,412	Examiner	: Teresa D. Wessendorf
Filed	: July 11, 2001	Docket	: 17027.003US1
Title	: ARTIFICIAL ANTIBODY POLYPEPTIDES		

APPEAL BRIEF

Mail Stop Appeal Brief - Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

The Final Office Action for this application was mailed August 09, 2006, and a Notice of Appeal and Pre-Appeal Brief Request for Review was mailed November 09, 2006. A Notice of Panel Decision from Pre-Appeal Brief Review was mailed on December 28, 2006, which Notice indicated that the application remains under appeal because there is at least one actual issue for appeal. Applicant respectfully appeals to the Board for review of the Examiner's final rejection.

(1)   Real Party in Interest.

The real party in interest is Research Corporation Technologies, Inc.

(2)   Related Appeals and Interferences.

There are no related appeals or interferences.

(3)   Status of Claims.

Claims 2, 3, 5, 6 and 9-53 have been canceled. Claims 1, 4, 7, 8 and 54-63 are pending and stand finally rejected. Applicant respectfully appeals the final rejection of claims 1, 4, 7, 8 and 54-63.

(4) **Status of Amendments.**

No amendments have been have been filed subsequent to the Final Office Action.

(5)   Summary of the Claimed Subject Matter.

The claimed subject matter relates to a modified fibronectin type III (Fn3) molecule comprising a stabilizing mutation of at least one residue involved in an unfavorable electrostatic interaction as compared to a wild-type Fn3, wherein the stabilizing mutation is a substitution of at least one of Asp 7, Asp 23 or Glu 9 with another amino acid residue (claim 1). The claimed subject matter also relates to a modified tenth type III module of fibronectin (FNfn10) molecule comprising a stabilizing mutation of at least one residue involved in an unfavorable electrostatic interaction as compared to a wild-type FNfn10 molecule, wherein the stabilizing mutation is a substitution of at least one of amino acid residues 7, 9 or 23 with another amino acid residue (claim 57). The claimed subject matter is described throughout the specification, for example, at page 6, lines 19-32; page 18, line 14 through page 20, line 5; page 35, line 4 through page 38, line 30; and at page 63, line 4 through page 77, line 23, and in the Figures referenced to in those sections.

(6)   Grounds of Rejection to be Reviewed on Appeal.

The issues being appealed are:

(A) whether claims 54-63 fail to comply with the written description requirement of 35 U.S.C. § 112, first paragraph by containing subject matter that was not described in the specification in such a way as reasonably convey to one skilled in the art that the inventor, at the time the application was filed, had possession of the claimed invention (a "new matter" rejection);

(B) whether claims 1, 8 and 54-63 fail to comply with the written description requirement of 35 U.S.C. § 112, first paragraph by containing subject matter that was not described in the specification in such a way as reasonably convey to one skilled in the art that the inventor, at the time the application was filed, had possession of the claimed invention; and

(C) whether claims 1, 4, 7-8, and 54-63 are unpatentable under 35 U.S.C. § 103(a) over Koide (WO 98/56915) or Lipovsek *et al.* (U.S. Patent No. 6,818,418) in view of Spector *et al.* (*Biochemistry*, *39*, 872-879 (2000)).

(7) Arguments

A. Claims 54-63 comply with the written description requirement of 35 U.S.C. § 112, first paragraph and do not contain "new matter."

The Examiner rejected claims 54-63 under 35 U.S.C. § 112, first paragraph, alleging that those claims fail to comply with the written description requirement. The Examiner alleges that those claims contain subject matter that was not described in the specification in such a way as to reasonably convey to one skilled in the art that the inventor, at the time the application was filed, had possession of the claimed invention. Specifically, the Examiner alleges that (1) the claimed "neutral" or "positively" charged amino acid residues and (2) the "open" amino acid residues at positions 7, 9 and 23 in claims 57-63 are not supported in the specification as filed. Applicant respectfully disagrees and submits that the originally-filed application provides sufficient support for those claims, *e.g.*, because the originally-filed application reasonably conveys to one having ordinary skill in the art that an Applicant had possession of the concepts of what is now claimed.

i. The "neutral" and/or "positively" charged amino acid residues recited in claims 54, 55, 56, 58, 59, 60 are supported in the specification as filed.

Independent claim 1 recites a modified fibronectin type III (Fn3) molecule comprising a stabilizing mutation of at least one residue involved in an unfavorable electrostatic interaction as compared to a wild-type Fn3, wherein the stabilizing mutation is a substitution of at least one of Asp 7, Asp 23 or Glu 9 with another amino acid residue. Claims 54-56 depend directly or indirectly from claim 1.

Independent claim 57 recites a modified tenth type III module of fibronectin (FNfn10) molecule comprising a stabilizing mutation of at least one residue involved in an unfavorable electrostatic interaction as compared to a wild-type FNfn10 molecule, wherein the stabilizing mutation is a substitution of at least one of amino acid residues 7, 9 or 23 with another amino acid residue. Claims 58-60 depend directly or indirectly from claim 57.

Thus, claims 54-56 are directed to modified Fn3 molecules that comprise a stabilizing mutation that is a substitution of at least one of Asp 7, Asp 23 or Glu 9 with a neutral or positively charged amino acid residue (claim 54) or with a neutral amino acid residue (claim 55) or with a positively charged amino acid residue (claim 56). Claims 58-60 are directed to modified FNfn10 molecules that comprise a stabilizing mutation that is a substitution of at least one of amino acid residues 7, 9 or 23 with a neutral or positively charged amino acid residue

(claim 58) or with a neutral amino acid residue (claim 59) or with a positively charged amino acid residue (claim 60).

The originally-filed disclosure provides sufficient support as long as it would have reasonably conveyed to one having ordinary skill in the art that an Applicant had possession of the concept of what is now claimed. *In re Anderson*, 176 U.S.P.Q. 331, 336 (C.C.P.A. 1973).

Applicant respectfully directs the Board's attention to the specification, for example, to page 71, lines 13-22, where Applicant explains the concept of substituting a neutral or positively charged amino acid residue for at least one of the negatively-charged residues Asp 7, Asp 23 or Glu 9 of an Fn3 molecule (FNfn10) to improve the stability of the molecule:

> "The spatial proximity of Asp 7 and 23, and Glu 9 explains the unfavorable electrostatic interactions in FNfn10 identified in this study. At low pH where these residues are protonated and neutral, the repulsive interactions are expected to be mostly relieved. Thus, it should be possible to improve the stability of FNfn10 at neutral pH, by removing the electrostatic repulsion between these three residues. Because Asp 7 is centrally located among the three residues, it was decided to mutate Asp 7. Two mutants, D7N and D7K were prepared. The former neutralizes the negative charge with a residue of virtually identical size. The latter places a positive charge at residue 7 and increases the size of the side chain."

Applicant respectfully submits that the originally-filed disclosure reasonably conveys to one having ordinary skill in the art that an Applicant had possession of the concept of what is now claimed, *i.e.*, the concept of substituting a neutral or positively charged amino acid residue for at least one of the negatively-charged residues Asp 7, Asp 23 or Glu 9 of an Fn3 molecule so as to improve the stability of the molecule.

The Examiner at page 4 of the Final Office Action alleges that the concept of substituting a neutral or positively charged amino acid residue for at least one of the negatively charged residues Asp 7, Asp 23 or Glu 9 so as to improve the stability of the molecule is not positive support for the numerous amino acid residues that are neutral or positively charged amino acid residues. However, even if for the sake of argument Applicant did not specifically list each of the other neutral or positively charged amino acid residues, Applicant submits that the originally-filed disclosure provides sufficient support for the claims because it reasonably conveys to one having ordinary skill in the art that an Applicant had possession of the concept of what is now claimed.

a. Claims 54-56

Claims 54-56 are directed to modified Fn3 molecules that comprise a stabilizing mutation that is a substitution of at least one of Asp 7, Asp 23 or Glu 9 with a neutral or positively charged amino acid residue (claim 54) or with a neutral amino acid residue (claim 55) or with a positively charged amino acid residue (claim 56). As described hereinabove, Applicant submits that the originally-filed disclosure provides sufficient support for claims 54-56 because it reasonably conveys to one having ordinary skill in the art that an Applicant had possession of the concept of what is now claimed, namely that at least one of the three specific amino acid residues is substituted with a neutral or positively charged amino acid residue that makes the Fn3 molecule more stable. Thus, the claims satisfy the written description requirement of 35 U.S.C. § 112, first paragraph.

b. Claims 58-60

Claims 58-60 are directed to modified FNfn10 molecules that comprise a stabilizing mutation that is a substitution of at least one of amino acid residues 7, 9 or 23 with a neutral or positively charged amino acid residue (claim 58) or with a neutral amino acid residue (claim 59) or with a positively charged amino acid residue (claim 60). Claims 58-60 are specifically directed to modified FNfn10 molecules. The sequences of FNfn10 molecules were well-known at the time the application was filed. Applicant taught which specific amino acid residues could be replaced (*i.e.*, residues 7, 9 and/or 23) in order to make the FNfn10 molecules more stable. Thus, Applicant respectfully submits that claims 58-60 are fully supported by originally-filed disclosure (*see* page 71, lines 13-22 of the specification) and, in view of what was known at the time the application was filed, satisfy the written description requirement of 35 U.S.C. § 112, first paragraph.

ii. The "open" amino acid residues at positions 7, 9 and 23 in claims 57-63 are supported in the specification as filed.

With respect to the amino acid residues at positions 7, 9 or 23 of claim 57, Applicant respectfully submits that the specification as-filed, *e.g.*, page 71, lines 13-22 (recited in Section 7(A)(i) above) and page 76, lines 6-12 (below), provides adequate support for the pending claims:

"The carboxyl triad (Asp 7 and 23, and Glu 9) is highly conserved in FNfn10 from nine different organisms that were available in the protein sequence databank at National Center for Biotechnology Information (www.ncbi.nlm.nih.gov). In these FNfn10 sequences, Asp 7 is conserved except one case where it is replaced with Asn, and Glu 9 is completely conserved. The position 23 is either Asp or Glu, preserving the negative charge. As was discovered in this study, the interactions among these residues are destabilizing."

Applicant notes that claims 57-63 are directed to modified FNfn10 molecules. Applicant respectfully submits that it is clear, e.g., from page 76, lines 6-12, that the originally-filed disclosure reasonably conveys to one having ordinary skill in the art that an Applicant had possession of the concept of what is now claimed, i.e., the concept of substituting an amino acid residue for at least one of amino acid residues 7, 9 or 23 of an FNfn10 molecule so as to improve the stability of the molecule.

a. Claims 57-60 and 63

As described hereinabove, Applicant respectfully submits that it is clear that the originally-filed disclosure reasonably conveys to one having ordinary skill in the art that Applicant had possession of the concept of what is now claimed in claims 57-60 and 63. Thus, these claims satisfy the written description requirement of 35 U.S.C. § 112, first paragraph.

b. Claim 61

Claim 61, which depends indirectly from claim 57, adds the further feature that the modified FNfn10 molecules at amino acid residues 7 or 23, or both, have been substituted with an asparagine (Asn) or lysine (Lys) residue. Applicant respectfully submits that claim 61 complies with the written description requirement of 35 U.S.C. § 112, first paragraph by only referring to amino acid residues 7 or 23, and specifying that the replacements are either Asn or Lys.

c. Claim 62

Claim 62, which depends indirectly from claim 57, is directed to modified FNfn10 molecules wherein amino acid residue 9 has been substituted with an asparagine (Asn) or lysine (Lys) residue. Applicant respectfully submits that claim 62 complies with the written description requirement of 35 U.S.C. § 112, first paragraph by only referring to only to amino acid residue 9, and specifying that the replacement is either Asn or Lys.

B. Claims 1, 8 and 54-63 comply with the written description requirement of 35 U.S.C. § 112, first paragraph.

The Examiner rejected claims 1, 8 and 54-63 under 35 U.S.C. § 112, first paragraph, alleging that those claims fail to comply with the written description requirement. The Examiner alleges that those claims contain subject matter that was not described in the specification in such a way as to reasonably convey to one skilled in the art that the inventor, at the time the application was filed, had possession of the claimed invention. The Examiner clarified the rejection at page 7 of the Final Office Action to indicate that the rejection is not based on the exclusion of inoperative embodiments of the invention.

Independent claims 1 and 57 are described hereinabove. Claims 8 and 54-56 depend directly or indirectly from claim 1. Claims 58-63 depend directly or indirectly from claim 57.

Applicant asserts that the specification as originally filed provides an adequate written description of the claimed invention. Applicant may show adequate written description by demonstrating that an invention is complete by disclosure of sufficiently detailed, relevant identifying characteristics that provide evidence that Applicant was in possession of the claimed invention, *i.e.*, complete or partial structure, other physical and/or chemical properties, functional characteristics when coupled with a known or disclosed correlation between function and structure, or some combination of such characteristics. *Enzo Biochem. v. Gen-Probe Inc.*, 323 F.3d 956, 963, 63 U.S.P.Q.2d 1609, 1613 (Fed. Cir. 2002). What is conventional or well known to one of ordinary skill in the art need not be disclosed in detail. *Hybritech Inc. v. Monoclonal Antibodies, Inc.*, 802 F.3d 1367, 1384, 231 U.S.P.Q. 81, 94 (Fed. Cir. 1986). Furthermore, the written description requirement states that the Applicant must describe the invention; it does not state that every invention must be described in the same way. As each field evolves, the balance also evolves between what is known and what is added by each inventive contribution. *Capon v. Eshhar v. Dudas*, 2005 U.S. App. LEXIS 16865 (Fed. Cir. 2005). Moreover, it is not necessary that every permutation within a generally operable invention be effective in order to obtain a generic claim, provided that the effect is sufficiently demonstrated to characterize a generic invention. *Capon v. Eshhar v. Dudas*, 2005 U.S. App. LEXIS 16865 (Fed. Cir. 2005).

Applicant provides structural characteristics of the claimed Fn3 molecules, including the claimed FNfn10 molecules. For example, the structure of wild-type Fn3 molecules are known (*see, e.g.*, Main *et al.* 1992, and page 18, line 14, through page 20, line 5 of the specification).

The claimed modified Fn3 molecules have a mutation of the Fn3 structure, *i.e.*, a substitution of at least one of amino acid residues 7, 9, or 23, *e.g.*, at least one of Asp 7, Asp 23 or Glu 9, with another amino acid residue. As such, Applicant has recited specific structural modifications of the Fn3 molecules. Thus, Applicant provides the art worker with structural characteristics of the claimed modified Fn3 molecules.

Applicant also provides functional characteristics of the claimed modified Fn3 molecules, namely, that the modified Fn3 molecules comprise a mutation that is a stabilizing mutation. A stabilizing mutation is defined in the specification at page 6, lines 20-24, as "a modification or change in the amino acid sequence of the Fn3 molecule, such as a substitution of one amino acid for another, that increases the melting point of the molecule by more than 0.1°C as compared to a molecule that is identical except for the change." Applicant provides a method for determining the melting point of the molecules in Example 19, which begins at page 63 of the specification. Thus, Applicant provides the art worker with functional characteristics of the claimed modified Fn3 molecules.

Applicant submits that the art worker is well apprised of amino acid residues to consider for substitution, including positive and neutral amino acids, and lists of those amino acids can be found in numerous sources. For example, Tables 3 and 4 of Chapter 2400 of the MPEP provide the art worker with amino acids (Table 3) and modified or unusual amino acids (Table 4) that could be considered for substitution for at least one of amino acid residues 7, 9 or 23 of the Fn3 molecule. In addition, the CRC Handbook of Chemistry and Physics also provides the art worker with information regarding specific properties of common amino acids (CRC Handbook of Chemistry and Physics; 76th Edition 1995-1996; CRC Press, Inc., Boca Raton, c1995, page 7-1; a copy of provided herewith). In particular, Applicant asserts that once Applicant discovered that amino acid residues 7, 9 and 23 of the Fn3 molecule were amino acids that contributed to unfavorable intra-molecular electrostatic interactions, one of ordinary skill in the art would know or be able to determine which amino acid residues could be substituted to enhance the stability of the Fn3. For example, Applicant submits that one of skill in the art would know that since both Asp and Glu have negative charges, the introduction of an amino acid that has either a neutral or positive charge would likely reduce or remove the unfavorable electrostatic interaction from amino acid residues 7, 9 and/or 23 and would thus provide a likely candidate for substitution. Applicant has provided the art worker evidence of this as a substitution of Asp 7 with a neutral

(*e.g.*, Asn) or positively-charged (*e.g.*, Lys) amino acid reduces the unfavorable interactions (page 75, lines 6-8 of the specification). And even if, for the sake of argument, the art worker lacked guidance as to which amino acid residue to select for substitution, the scope of the claims is described functionally as the substitution is recited to stabilize the molecule. As such, the art worker would need only to test the substitution(s) at the recited position(s) to determine whether the substitution stabilized the molecule using, *e.g.*, the assay described in Example 19. Applicant submits that such testing would not be undue.

Thus, Applicant has provided structural characteristics of the claimed modified Fn3 molecules as the structure of wild-type Fn3 molecules were known to the art worker at the time the application was filed. Applicant has recited specific structural modifications to the known Fn3 molecule, *i.e.*, the modified Fn3 molecule has a substitution of at least one of amino acid residues 7, 9 or 23, *e.g.*, Asp 7, Asp 23 or Glu 9. Applicant has also recited functional characteristics of the claimed modified Fn3 molecules, namely, that the modified Fn3 molecules comprise a stabilizing mutation, which mutation is functionally described in the specification together with an assay to measure the functional characteristic. Applicant has further provided examples of stabilizing mutations of the recited amino acids. Thus, it is respectfully asserted that Applicant has provided adequate written description of the claimed modified Fn3 molecules as Applicant has disclosed in sufficient detail the relevant identifying structural and functional characteristics that provide evidence that the Applicant was in possession of the full scope of the claimed invention at the time the application was filed. Thus, Applicant submits that the claims satisfy the written description requirements of 35 U.S.C. § 112, first paragraph and requests that the Board withdraw this rejection of the claims.

i. Claims 1 and 8

As described hereinabove, Applicant submits that claims 1 and 8 satisfy the written description requirements of 35 U.S.C. § 112, first paragraph.

ii. Claims 54-56

Claims 54-56 depend directly or indirectly from claim 1. Applicant submits that claim 1 satisfies the written description requirements of 35 U.S.C. § 112, first paragraph. Claims 54-56 further define the invention and are directed to modified Fn3 molecules that comprise a stabilizing mutation that is a substitution of at least one of Asp 7, Asp 23 or Glu 9 with a neutral or positively charged amino acid residue (claim 54) or with a neutral amino acid residue (claim

55) or with a positively charged amino acid residue (claim 56). Thus, claims 54-56 provide additional structural characteristics of the claimed modified Fn3 molecules and satisfy the written description requirement of 35 U.S.C. § 112, first paragraph.

iii. Claim 57

Claim 57 is directed to FNfn10 molecules, which wild-type molecules are a subset of the wild-type Fn3 molecules of claim 1. Thus, Applicant submits that claim 57 satisfies the written description requirements of 35 U.S.C. § 112, first paragraph by providing additional structural characteristics of the claimed modified Fn3 molecules because FNfn10 molecules were known at the time the application was filed, and Applicant provided an adequate description of which amino acid residues to modify.

iv. Claims 58-60

Claims 58-60 depend directly or indirectly from claim 57. Applicant submits that claim 57 satisfies the written description requirements of 35 U.S.C. § 112, first paragraph. Claims 58-60 further define the invention and are directed to modified FNfn10 molecules that comprise a stabilizing mutation that is a substitution of at least one of amino acids 7, 9 or 23 with a neutral or positively charged amino acid residue (claim 58) or with a neutral amino acid residue (claim 59) or with a positively charged amino acid residue (claim 60). Thus, claims 54-56 provide additional structural characteristics of the claimed modified Fn3 molecules and satisfy the written description requirement of 35 U.S.C. § 112, first paragraph.

v. Claims 61 and 62

Claims 61 and 62 depend from claim 58, which depends from claim 57. Applicant submits that claims 57 and 58 satisfy the written description requirements of 35 U.S.C. § 112, first paragraph. Claims 61 and 62 further define the invention and are directed to modified FNfn10 molecules that comprise a stabilizing mutation that is a substitution of at least one of amino acids 7, 9 or 23 with a neutral or positively charged amino acid residue, wherein amino acid residues 7 or 23, or both, have been substituted with an asparagine (Asn) or lysine (Lys) residue (claim 61) or wherein amino acid residue 9 has been substituted with an asparagine (Asn) or lysine (Lys) residue (claim 62). Thus, claims 61 and 62 provide additional structural characteristics of the claimed molecules and satisfy the written description requirement of 35 U.S.C. § 112, first paragraph.

C. Claims 1, 8 and 54-63 are patentable over Koide, Lipovsek and/or Spector.

The Examiner rejected claims 1, 4, 7-8 and 54-63 under 35 U.S.C. § 103(a), alleging that those claims are unpatentable over Koide (WO 98/56915; hereinafter Koide) or Lipovsek et al. (U.S. Patent No. 6,818,418; hereinafter Lipovsek) in view of Spector et al. (Biochemistry, 39, 872-879 (2000); hereinafter Spector).

Claims 1 and 57 are independent claims. Claims 8 and 54-56 depend directly or indirectly from claim 1. Claims 58-63 depend directly or indirectly from claim 57. All of these claims are described hereinabove.

Applicant respectfully submits that the Examiner has not demonstrated that the claims are prima facie obvious in view of the cited documents, for example, because the Examiner has not established that the cited documents teach or suggest all the claim limitations. And, even if, for the sake or argument, the cited documents teach or suggests all the claim limitations, Applicant respectfully submits that the Examiner has not established the suggestion or motivation, either in the cited documents themselves or in the knowledge generally available to an art worker, to modify the documents or to combine document teachings so as to arrive at the claimed invention. Further, Applicant respectfully submits that the Examiner is improperly relying on an "obvious to try" standard.

Koide relates to Fn3 polypeptide monobodies. Only mutant fibronectin molecules with reduced stability relative to wild type fibronectin are disclosed in Koide (e.g., Figure 16 and Example XVII).

Lipovsek relates to antibody mimics that are based on the structure of an Fn3 (column 7, lines 63-65). Lipovsek states that for the human [10]Fn3 sequence, at a minimum, amino acids 1-9, 44-50, 61-54, 82-94 (edges of beta sheets); 19, 21, 30-46 (even), 79-65 (odd) (solvent-accessible faces of both beta sheets); 21-31, 51-56, 76-88 (CDR-like solvent-accessible loops); and 14-16 and 36-45 (other solvent-accessible loops and beta turns) may be randomized to evolve new or improved compound-binding proteins (column 9, lines 24-31).

Spector relates to the electrostatic contributions that charged and polar side chains make on the overall stability of a 41-residue protein (first sentence of the Abstract), a protein that is based on the peripheral subunit-binding domain, derived from the dihydrolipoamide acetyltransferase component of the pyruvate dehydrogenase multienzyme complex from Bacillus stearothermophilus (page 873, first column, second full paragraph).

A rejection of obviousness under 35 U.S.C. § 103 requires that the Examiner establish a *prima facie* case of obviousness. To establish a *prima facie* case of obviousness, the Examiner has the burden to establish three basic elements. First, the Examiner must establish that there is some suggestion or motivation, either in the cited documents themselves or in the knowledge generally available to an art worker, to modify the documents or to combine document teachings so as to arrive at the claimed invention. Second, the Examiner must establish that there is a reasonable expectation of success. Finally, the Examiner must establish that the prior art documents teach or suggests all the claim limitations. M.P.E.P. § 2143.

At page 11 of the Office Action, the Examiner states neither Koide nor Lipovsek teaches that the regions of Fn3 containing amino acids 7, 9 or 23 are involved in an unfavorable electrostatic interaction. Applicant respectfully submits that Spector does not remedy the deficiencies of Koide and Lipovsek because Spector does not teach or suggest that the regions of Fn3 containing amino acids 7, 9 or 23 are involved in an unfavorable electrostatic interaction. Spector is related to the peripheral subunit-binding domain, derived from the dihydrolipoamide acetyltransferase component of the pyruvate dehydrogenase multienzyme complex from *Bacillus stearothermophilus*, not to Fn3. Thus, Applicant submits that the Examiner has not established that the cited documents teach or suggest all the claim limitations, *e.g.*, a modified Fn3 or FNfn10 molecule comprising a stabilizing mutation of at least one residue involved in an unfavorable electrostatic interaction as compared to a wild-type Fn3 or FNfn10 molecule, wherein the stabilizing mutation is a substitution of at least one of amino acid residues 7, 9 or 23 (*e.g.*, Asp 7, Asp 23 or Glu 9) with another amino acid residue.

Applicant submits that the Examiner has not established the suggestion or motivation, either in the cited documents themselves or in the knowledge generally available to an art worker, to modify the documents or to combine document teachings so as to arrive at the claimed invention. At pages 11-12 of the Final Office Action, the Examiner alleges that it would have been obvious to one having ordinary skill in the art at the time the invention was made to determine <u>whether</u> the amino acids in the 1-9 or 21-31 regions of Fn3 of Koide or Lipovsek are involved in an unfavorable electrostatic interaction as taught by Spector (underline added). Lipovsek states that for the human [10]Fn3 sequence, at a minimum, amino acids 1-9, 44-50, 61-54, 82-94 (edges of beta sheets); 19, 21, 30-46 (even), 79-65 (odd) (solvent-accessible faces of both beta sheets); 21-31, 51-56, 76-88 (CDR-like solvent-accessible loops); and 14-16 and 36-45

(other solvent-accessible loops and beta turns) may be randomized to evolve new or improved compound-binding proteins (column 9, lines 24-31). However, as stated in M.P.E.P. § 2145(X)(B), "obvious to try" is not the standard under 35 U.S.C. § 103. Specifically, trying each of numerous possible choices until one possibly arrived at a successful result, where the prior art gave either no indication of which parameters were critical or no direction as to which of many possible choices is likely to be successful, is an improper "obvious to try" standard. Applicant respectfully submits that the Examiner is improperly relying on an "obvious to try" standard by suggesting that the art worker could have tried each of numerous possible choices, *i.e.*, the listing of amino acids, until the art worker possibly arrived at a successful result.

Thus, Applicant respectfully submits that the cited documents, neither alone nor in combination, teach a modified Fn3 molecule comprising a stabilizing mutation of at least one residue involved in an unfavorable electrostatic interaction as compared to a wild-type Fn3, wherein the stabilizing mutation is a substitution of at least one of Asp 7, Asp 23 or Glu 9 with another amino acid residue. Nor do the cited documents, either alone or in combination, teach a FNfn10 molecule comprising a stabilizing mutation of at least one residue involved in an unfavorable electrostatic interaction as compared to a wild-type FNfn10 molecule, wherein the stabilizing mutation is a substitution of at least one of amino acid residues 7, 9 or 23 with another amino acid residue. Thus, Applicant respectfully requests that the Board withdraw the rejection of the claims under 35 U.S.C. § 103(a).

Each claim is argued separately.

Applicant respectfully submits that the Examiner has not separately demonstrated that any of claims 1, 4, 7-8 or 54-63 are separately *prima facie* obvious in view of the cited documents, for example, because the Examiner has not established that the cited documents teach or suggest the claim limitation of each separate claim. And, even if, for the sake or argument, the cited documents teach or suggests all the claim limitations, Applicant respectfully submits that the Examiner has not established the suggestion or motivation, either in the cited documents themselves or in the knowledge generally available to an art worker, to modify the documents or to combine document teachings so as to arrive at the claimed invention of each separate claim. Because of the specific elements of each claim, each claim is argued separately.

1. Claim 1

As described hereinabove, Applicant respectfully submits that claim 1, which is directed to modified Fn3 molecules comprising a stabilizing mutation of at least one residue involved in an unfavorable electrostatic interaction as compared to a wild-type Fn3, wherein the stabilizing mutation is a substitution of at least one of Asp 7, Asp 23 or Glu 9 with another amino acid residue, is patentable over Koide, Lipovsek and/or Spector.

2. Claim 4

Claim 4 depends from claim 1 and specifically recites that for the modified Fn3 molecule, Asp 7 or Asp 23, or both, have been substituted with an asparagine (Asn) or lysine (Lys) residue. Applicant submits that the Examiner has not demonstrated that any of Koide, Lipovsek and/or Spector teach such an element, nor has the Examiner established the suggestion or motivation to modify the documents or to combine document teachings so as to arrive such a modified Fn3 molecule.

3. Claim 7

Claim 7 depends from claim 1 and specifically recites that for the modified Fn3 molecule, Glu 9 has been substituted with an asparagine (Asn) or lysine (Lys) residue. Applicant submits that the Examiner has not demonstrated that any of Koide, Lipovsek and/or Spector teach such an element, nor has the Examiner established the suggestion or motivation to modify the documents or to combine document teachings so as to arrive such a modified Fn3 molecule.

4. Claim 8

Claim 8 depends from claim 1 and specifically recites that for the modified Fn3 molecule, Asp 7, Asp 23, and Glu 9 have been substituted with at least one other amino acid residue. Applicant submits that the Examiner has not demonstrated that any of Koide, Lipovsek and/or Spector teach such an element, nor has the Examiner established the suggestion or motivation to modify the documents or to combine document teachings so as to arrive such a modified Fn3 molecule.

5. Claim 54

Claim 54 depends from claim 1 and specifically recites that for the modified Fn3 molecule, the stabilizing mutation is a substitution of at least one of Asp 7, Asp 23 or Glu 9 with a neutral or positively charged amino acid residue. Applicant submits that the Examiner has not demonstrated that any of Koide, Lipovsek and/or Spector teach such an element, nor has the

Examiner established the suggestion or motivation to modify the documents or to combine document teachings so as to arrive such a modified Fn3 molecule.

6. Claim 55

Claim 55 depends from claim 54 and specifically recites that for the modified Fn3 molecule, the stabilizing mutation is a substitution of at least one of Asp 7, Asp 23 or Glu 9 with a neutral amino acid residue. Applicant submits that the Examiner has not demonstrated that any of Koide, Lipovsek and/or Spector teach such an element, nor has the Examiner established the suggestion or motivation to modify the documents or to combine document teachings so as to arrive such a modified Fn3 molecule.

7. Claim 56

Claim 56 depends from claim 54 and specifically recites that for the modified Fn3 molecule, the stabilizing mutation is a substitution of at least one of Asp 7, Asp 23 or Glu 9 with a positively charged amino acid residue. Applicant submits that the Examiner has not demonstrated that any of Koide, Lipovsek and/or Spector teach such an element, nor has the Examiner established the suggestion or motivation to modify the documents or to combine document teachings so as to arrive such a modified Fn3 molecule.

8. Claim 57

As described hereinabove, Applicant respectfully submits that claim 57, which is directed to modified FNfn10 molecules comprising a stabilizing mutation of at least one residue involved in an unfavorable electrostatic interaction as compared to a wild-type FNfn10 molecule, wherein the stabilizing mutation is a substitution of at least one of amino acid residues 7, 9 or 23 with another amino acid residue, is patentable over Koide, Lipovsek and/or Spector.

9. Claim 58

Claim 58 depends from claim 57 and specifically recites that for the modified FNfn10 molecule, the stabilizing mutation is a substitution of at least one of amino acid residues 7, 9 or 23 with a neutral or positively charged amino acid residue. Applicant submits that the Examiner has not demonstrated that any of Koide, Lipovsek and/or Spector teach such an element, nor has the Examiner established the suggestion or motivation to modify the documents or to combine document teachings so as to arrive such a modified FNfn10 molecule.

10. Claim 59

Claim 59 depends from claim 58 and specifically recites that for the modified FNfn10 molecule, the stabilizing mutation is a substitution of at least one of amino acid residues 7, 9 or 23 with a neutral amino acid residue. Applicant submits that the Examiner has not demonstrated that any of Koide, Lipovsek and/or Spector teach such an element, nor has the Examiner established the suggestion or motivation to modify the documents or to combine document teachings so as to arrive such a modified FNfn10 molecule.

11. Claim 60

Claim 60 depends from claim 58 and specifically recites that for the modified FNfn10 molecule, the stabilizing mutation is a substitution of at least one of amino acid residues 7, 9 or 23 with a positively charged amino acid residue. Applicant submits that the Examiner has not demonstrated that any of Koide, Lipovsek and/or Spector teach such an element, nor has the Examiner established the suggestion or motivation to modify the documents or to combine document teachings so as to arrive such a modified FNfn10 molecule.

12. Claim 61

Claim 61 depends from claim 58 and specifically recites that for the modified FNfn10 molecule, amino acid residues 7 or 23, or both, have been substituted with an asparagine (Asn) or lysine (Lys) residue. Applicant submits that the Examiner has not demonstrated that any of Koide, Lipovsek and/or Spector teach such an element, nor has the Examiner established the suggestion or motivation to modify the documents or to combine document teachings so as to arrive such a modified FNfn10 molecule.

13. Claim 62

Claim 62 depends from claim 58 and specifically recites that for the modified FNfn10 molecule, amino acid residue 9 has been substituted with an asparagine (Asn) or lysine (Lys) residue. Applicant submits that the Examiner has not demonstrated that any of Koide, Lipovsek and/or Spector teach such an element, nor has the Examiner established the suggestion or motivation to modify the documents or to combine document teachings so as to arrive such a modified FNfn10 molecule.

14. Claim 63

Claim 63 depends from claim 57 and specifically recites that for the modified FNfn10 molecule, amino acid residues 7, 9 and 23 have been substituted with at least one other amino

acid residue. Applicant submits that the Examiner has not demonstrated that any of Koide, Lipovsek and/or Spector teach such an element, nor has the Examiner established the suggestion or motivation to modify the documents or to combine document teachings so as to arrive such a modified FNfn10 molecule.

At page 2 of the Final Office Action, the Examiner objected to a hyperlink in the specification. Applicant will gladly amend the objected-to paragraph to correct the hyperlink upon notification of allowable claims.

Applicant respectfully submits that the claims are in condition for allowance, and notification to that effect is respectfully requested. If necessary, please charge any additional fees or credit overpayment to Deposit Account 50-3503.

Respectfully submitted,
Shohei Koide
By his Representatives,
Viksnins Harris & Padys PLLP
PO Box 111098
St. Paul, MN 55111-1098
(952) 876-4094

Date: *March 28, 2007*
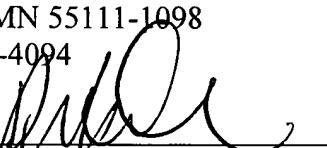
By: _____
Peter L. Malen
Reg. No. 44,894

CERTIFICATE OF MAILING BY FIRST CLASS MAIL

I hereby certify under 37 CFR §1.8(a) that this correspondence is being deposited with the United States Postal Service as first class mail with sufficient postage on the date indicated below and is addressed to the Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

March 28, 2007
_____
Date of Deposit

_____
Signature
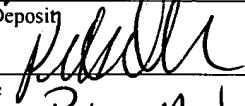
Peter Malon

Typed or Printed Name of Person Signing Certificate

(8) Claims Appendix.

1. A modified fibronectin type III (Fn3) molecule comprising a stabilizing mutation of at least one residue involved in an unfavorable electrostatic interaction as compared to a wild-type Fn3, wherein the stabilizing mutation is a substitution of at least one of Asp 7, Asp 23 or Glu 9 with another amino acid residue.

4. The Fn3 of claim 1, wherein Asp 7 or Asp 23, or both, have been substituted with an asparagine (Asn) or lysine (Lys) residue.

7. The Fn3 of claim 1, wherein Glu 9 has been substituted with an asparagine (Asn) or lysine (Lys) residue.

8. The Fn3 of claim 1, wherein Asp 7, Asp 23, and Glu 9 have been substituted with at least one other amino acid residue.

54. The Fn3 of claim 1, wherein the stabilizing mutation is a substitution of at least one of Asp 7, Asp 23 or Glu 9 with a neutral or positively charged amino acid residue.

55. The Fn3 of claim 54, wherein the stabilizing mutation is a substitution of at least one of Asp 7, Asp 23 or Glu 9 with a neutral amino acid residue.

56. The Fn3 of claim 54, wherein the stabilizing mutation is a substitution of at least one of Asp 7, Asp 23 or Glu 9 with a positively charged amino acid residue.

57. A modified tenth type III module of fibronectin (FNfn10) molecule comprising a stabilizing mutation of at least one residue involved in an unfavorable electrostatic interaction as compared to a wild-type FNfn10 molecule, wherein the stabilizing mutation is a substitution of at least one of amino acid residues 7, 9 or 23 with another amino acid residue.

58.   The modified FNfn10 of claim 57, wherein the stabilizing mutation is a substitution of at least one of amino acid residues 7, 9 or 23 with a neutral or positively charged amino acid residue.

59.   The modified FNfn10 of claim 58, wherein the stabilizing mutation is a substitution of at least one of amino acid residues 7, 9 or 23 with a neutral amino acid residue.

60.   The modified FNfn10 of claim 58, wherein the stabilizing mutation is a substitution of at least one of amino acid residues 7, 9 or 23 with a positively charged amino acid residue.

61.   The modified FNfn10 of claim 58, wherein amino acid residues 7 or 23, or both, have been substituted with an asparagine (Asn) or lysine (Lys) residue.

62.   The modified FNfn10 of claim 58, wherein amino acid residue 9 has been substituted with an asparagine (Asn) or lysine (Lys) residue.

63.   The modified FNfn10 of claim 57, wherein amino acid residues 7, 9 and 23 have been substituted with at least one other amino acid residue.

(9)   Evidence Appendix.

A. Manual of Patent Examining Procedure, Tables 3 and 4 (1998).

   Please refer to the Amendment and Response mailed on September 21, 2005.

B. CRC Handbook of Chemistry and Physics; 76th Edition 1995-1996; CRC Press, Inc., Boca Raton, c1995, page 7-1.

   Please refer to the Amendment and Response mailed on September 21, 2005.

C. Main *et al.*, "The three-dimensional structure of the tenth type III module of fibronectin: An insight into RGD-mediated interactions", *Cell, 71*, 671-678 (1992).

   Please refer to the Information Disclosure Statement mailed on July 11, 2001.

D. WO 98/56915

   Please refer to the Information Disclosure Statement mailed on April 11, 2002.

E. U.S. Patent No. 6,818,418

   Please refer to the Information Disclosure Statement mailed on July 11, 2001.

F. Spector *et al.* "Rational modification of protein stability by the mutation of charged surface residues", *Biochemistry, 39*, 872-879 (2000).

   Please refer to the Information Disclosure Statement mailed on September 21, 2005.

(10)   Related Proceedings Appendix.

There have been no decisions rendered by a court or the Board in the appeal of Application Serial No. 09/903,412.

Table 3: List of Amino Acids

Symbol	Meaning
Ala	Alanine
Cys	Cysteine
Asp	Aspartic Acid
Glu	Glutamic Acid
Phe	Phenylalanine
Gly	Glycine
His	Histidine
Ile	Isoleucine
Lys	Lysine
Leu	Leucine
Met	Methionine
Asn	Asparagine
Pro	Proline
Gln	Glutamine
Arg	Arginine
Ser	Serine
Thr	Threonine
Val	Valine
Trp	Tryptophan
Tyr	Tyrosine
Asx	Asp or Asn
Glx	Glu or Gln
Xaa	unknown or other

WIPO Standard ST.25 (1998), Appendix 2, Table 4, provides that modified and unusual amino acids may be represented as the corresponding unmodified amino acids in the sequence itself if the modified or unusual amino acid is one of those listed below and the modification is further described in the Feature section of the Sequence Listing. The codes from the list below may be used in the description (i.e., the specification and drawings, or in Sequence Listing) but these codes may not be used in the sequence itself.

Table 4: List of Modified and Unusual Amino Acids

Symbol	Meaning
Aad	2-Aminoadipic acid
bAad	3-Aminoadipic acid
bAla	beta-Alanine, beta-Aminopropionic acid
Abu	2-Aminobutyric acid
4Abu	4-Aminobutyric acid, piperidinic acid
Acp	6-Aminocaproic acid
Ahe	2-Aminoheptanoic acid
Aib	2-Aminoisobutyric acid
bAib	3-Aminoisobutyric acid
Apm	2-Aminopimelic acid
Dbu	2,4-Diaminobutyric acid
Des	Desmosine
Dpm	2,2'-Diaminopimelic acid
Dpr	2,3-Diaminopropionic acid
EtGly	N-Ethylglycine
EtAsn	N-Ethylasparagine
Hyl	Hydroxylysine
aHyl	allo-Hydroxylysine
3Hyp	3-Hydroxyproline
4Hyp	4-Hydroxyproline
Ide	Isodesmosine
aIle	allo-Isoleucine
MeGly	N-Methylglycine, sarcosine
MeIle	N-Methylisoleucine
MeLys	6-N-Methyllysine
MeVal	N-Methylvaline
Nva	Norvaline
Nle	Norleucine

Orn	Ornithine

WIPO Standard ST.25 (1998), Appendix 2, Table 5, provides for feature keys related to DNA sequences.

PROPERTIES OF COMMON AMINO ACIDS

This table gives selected properties of 20 α-amino acids commonly found in proteins. The structures of these amino acids are given in a separate table. The compounds are listed in alphabetical order by the three-letter symbols. Dissociation constants refer to aqueous solutions at 25° C.

M_r — Molecular weight
T_m — Melting point
pK_a — Negative of the logarithm of the dissociation constant for the α-COOH group
pK_b — Negative of the logarithm of the dissociation constant for the α-NH$_3^+$ group
pK_x — Negative of the logarithm of the dissociation constant for any other group present in the molecule
pI — pH at the isoelectronic point
S — Solubility in water at 25° C in units of grams per kilogram of water

Symbol	Name	Mol. form.	M_r	$T_m/°C$	pK_a	pK_b	pK_x	pI	S
Ala	Alanine	$C_3H_7NO_2$	89.09	297	2.34	9.69		6.00	167
Arg	Arginine	$C_6H_{14}N_4O_2$	174.20	238	2.17	9.04	12.48	10.76	181
Asn	Asparagine	$C_4H_8N_2O_3$	132.12	236	2.02	8.80		5.41	25
Asp	Aspartic acid	$C_4H_7NO_4$	133.10	270	1.88	9.60	3.65	2.77	5
Cys	Cysteine	$C_3H_7NO_2S$	121.16	178	1.96	10.28	8.18	5.07	
Gln	Glutamine	$C_5H_{10}N_2O_3$	146.15	185	2.17	9.13		5.65	42
Glu	Glutamic acid	$C_5H_9NO_4$	147.13	249	2.19	9.67	4.25	3.22	
Gly	Glycine	$C_2H_5NO_2$	75.07	290	2.34	9.60		5.97	251
His	Histidine	$C_6H_9N_3O_2$	155.16	277	1.82	9.17	6.00	7.59	43
Ile	Isoleucine	$C_6H_{13}NO_2$	131.17	284	2.36	9.60		6.02	34
Leu	Leucine	$C_6H_{13}NO_2$	131.17	337	2.36	9.60		5.98	23
Lys	Lysine	$C_6H_{14}N_2O_2$	146.19	224—225	2.18	8.95	10.53	9.74	6
Met	Methionine	$C_5H_{11}NO_2S$	149.21	283	2.28	9.21		5.74	56
Phe	Phenylalanine	$C_9H_{11}NO_2$	165.19	284	1.83	9.13		5.48	29
Pro	Proline	$C_5H_9NO_2$	115.13	222	1.99	10.60		6.30	1622
Ser	Serine	$C_3H_7NO_3$	105.09	228	2.21	9.15		5.68	422
Thr	Threonine	$C_4H_9NO_3$	119.12	253	2.09	9.10		5.60	97
Trp	Tryptophan	$C_{11}H_{12}N_2O_2$	204.23	282	2.83	9.39		5.89	12
Tyr	Tyrosine	$C_9H_{11}NO_3$	181.19	344	2.20	9.11	10.07	5.66	0.5
Val	Valine	$C_5H_{11}NO_2$	117.15	292-295	2.32	9.62		5.96	58

REFERENCES

G. D. Fasman, Ed., *Practical Handbook of Biochemistry and Molecular Biology*, CRC Press, Boca Raton, FL, 1989.
E. L. Smith, et al., *Principles of Biochemistry*, 7th ed., McGraw Hill, New York, 1983.
H. J. Hinz, Ed., *Thermodynamic Data for Biochemistry and Biotechnology*, Springer-Verlag, Heidelberg, 1986.

The Three-Dimensional Structure of the Tenth Type III Module of Fibronectin: An Insight into RGD-Mediated Interactions

Alison L. Main,* Timothy S. Harvey,* Martin Baron,†
Jonathan Boyd,* and Iain D. Campbell*
*Department of Biochemistry
University of Oxford
South Parks Road
Oxford OX1 3QU
England

Summary

The solution structure of the tenth type III module of fibronectin has been determined using nuclear magnetic resonance techniques. The molecule has a fold similar to that of immunoglobulin domains, with seven β strands forming two antiparallel β sheets, which pack against each other. Both β sheets contribute conserved hydrophobic residues to a compact core. The topology is more similar to that of domain 2 of CD4, PapD, and the extracellular domain of the human growth hormone receptor than to that of immunoglobulin C domains. The module contains an Arg-Gly-Asp sequence known to be involved in cell adhesion. This tripeptide is solvent exposed and lies on a conformationally mobile loop between strands F and G, consistent with its cell adhesion function.

Introduction

Fibronectin is a multifunctional protein found in the extracellular matrix and serum. Its diverse biological roles rely on an ability to bind to components of the extracellular matrix and receptors on the cell surface (Hynes, 1990). It is composed of three different kinds of structural unit, referred to as types I, II, and III (Ruoslahti, 1988). Solution structures of type I and type II modules have been determined using nuclear magnetic resonance (NMR) techniques (Baron et al., 1990; Constantine et al., 1992); the structure of the tenth type III module is described below. The type III module is characterized by a consensus sequence (Patthy, 1991; Baron et al., 1991) of approximately 90 amino acids. Over 140 occurrences of this module have been found in a wide range of proteins including extracellular cell adhesion molecules (Patthy, 1991; Mayford et al., 1992) and intracellular proteins involved in muscle filament formation (Labeit et al., 1990). Knowledge of its structure should thus prove to be a useful tool in the modeling of a wide range of proteins.

The cell adhesion activity of fibronectin has been localized to an Arg-Gly-Asp (RGD) sequence lying close to the C-terminus of of the tenth type III module (Pierschbacher and Ruoslahti, 1984). RGD sequences have also been found to be responsible for the cell adhesive properties

†Present address: Boyer Center for Molecular Medicine, Yale University School of Medicine, 295 Congress Avenue, New Haven, Connecticut 06536.

of a number of other proteins, including fibrinogen, von Willebrand factor, and vitronectin (Hynes, 1992). All known RGD receptors are members of the integrin family of cell adhesion molecules; however, the mechanism and specificity of integrin binding to RGD-containing ligands remain unclear. Recent studies have shown that regions of fibronectin other than the RGD sequence are necessary for full adhesive activity (Obara et al., 1988; Aota et al., 1991; Nagai et al., 1991), but it remains unclear whether such sequences stabilize the conformation of the RGD sequence or provide additional sites for interaction with the integrin (Ruoslahti and Pierschbacher, 1987; Mosher, 1989; Yamada, 1991). Short RGD-containing peptides have been shown to mimic a number of the properties of cell adhesive proteins, with differing conformations of the RGD motif resulting in changes in binding activity and integrin specificity (D'Souza et al., 1991). Consequently, knowledge of the conformation of the RGD sequence in the context of the type III module is of considerable value.

We have previously described NMR studies determining the secondary structure of the module, produced by heterologous gene expression in yeast (Baron et al., 1992). The module produced was shown to have cell binding activity, suggesting that correct folding had occurred. In this paper we describe the overall tertiary fold and dynamic properties of this module and the characteristics of the RGD motif it contains.

Results and Discussion

The experimental data from which three-dimensional structures were derived comprised 1084 nuclear Overhauser enhancements (NOEs) (735 of these were judged to be structurally significant NOE distance restraints, using the program DIANA [Güntert et al., 1991]), 66 hydrogen bond restraints, and 117 dihedral angle restraints (71 ϕ, 26 ψ, and 20 χ_1). The NOE restraints are unevenly distributed, with some parts of the molecule lacking long-range restraints. Consequently, some regions of the structure are much better defined than others. The initial stage of the structure calculation generated 45 structures, from which 36 structures were selected on the basis of NOE and restrained dihedral angle energies. An overlay of the 36 structures is shown in Figure 1a. A MOLSCRIPT diagram (Kraulis, 1991) labeling the seven β strands is shown in Figure 1b. A summary of energy terms and deviations from idealized geometry is given in Table 1.

Structure of the Type III Module

The structure of the type III module consists of seven β strands, which form a sandwich of two antiparallel β sheets, one containing three strands (ABE) and the other four strands (C'CFG). The triple-stranded β sheet consists of residues Glu-9–Thr-14 (A), Ser-17–Asp-23 (B), and Thr-56–Ser-60 (E). Location of the secondary structure elements was carried out using Quanta software (Polygen Corp.), which locates secondary structure elements on the
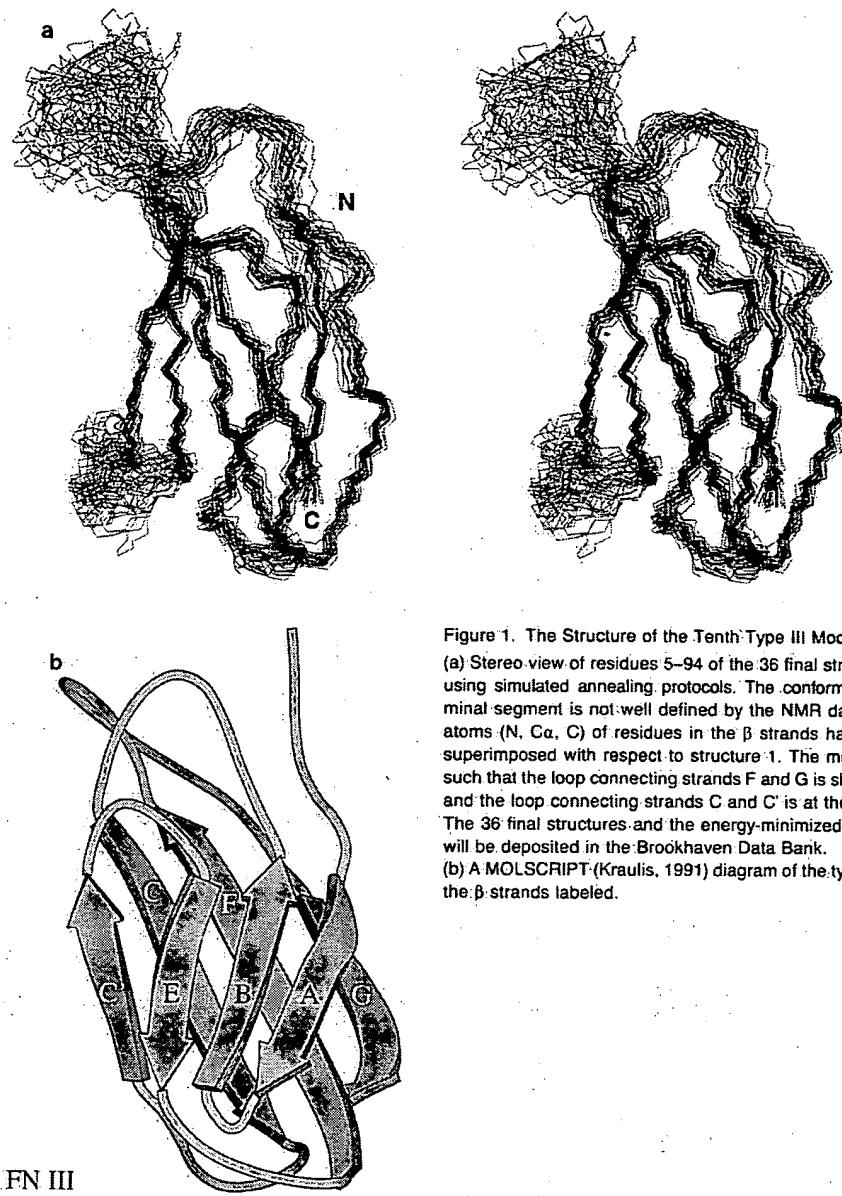
Figure 1. The Structure of the Tenth Type III Module of Fibronectin

(a) Stereo view of residues 5-94 of the 36 final structures calculated using simulated annealing protocols. The conformation of the N-terminal segment is not well defined by the NMR data. The backbone atoms (N, Cα, C) of residues in the β strands have been optimally superimposed with respect to structure 1. The molecule is oriented such that the loop connecting strands F and G is shown at the top left and the loop connecting strands C and C' is at the lower left corner. The 36 final structures and the energy-minimized average structure will be deposited in the Brookhaven Data Bank.

(b) A MOLSCRIPT (Kraulis, 1991) diagram of the type III module, with the β strands labeled.

basis of backbone dihedral angles and hydrogen bonds using the DSSP algorithm (Kabsch and Sander, 1983). There is a "classic" β bulge involving residues Val-11, Ala-12, and Leu-19 (Richardson et al., 1978). This bulge occurs between a pair of closely spaced hydrogen bonds, formed from both HN and CO of Leu-19, with Val-11 in approximately α-helical conformation (ϕ −80°, ψ −45°) and Ala-12 in approximately normal β sheet conformation (ϕ −160°, ψ +165°). The turn between strands A and B is well defined and corresponds to a 2:2 turn, which appears to be a distorted type I β turn, with average ϕ, ψ angles of −60°, +5° for residue i + 1 and −160°, +15° for residue i + 2 (Wilmot and Thornton, 1990). The four-stranded β sheet consists of residues Tyr-31–Glu-38 (C), Gln-46–Pro-51 (C'), Val-66–Thr-76 (F), and Ile-88–Thr-94 (G). The loops between

strands C and C' and strands F and G are 9:9 and 13:13 turns, respectively (Sibanda et al., 1989). Both β sheets have a right-handed twist and they stack on top of each other to enclose a hydrophobic core. Figure 2 shows the structure of the module, highlighting the secondary structure elements and the positions of Arg-78, Gly-79, and Asp-80 at the apex of the F–G loop.

Having determined the structure of this module, it is possible to address the significance of the highly conserved residues in the type III family. An alignment of the type III modules of fibronectin is shown in Figure 3. The majority of the conserved residues contribute to the hydrophobic core, with the invariant hydrophobic residues Trp-22 and Tyr-68 lying toward the N-terminal and C-terminal ends of the core, respectively. Other module

Table 1. Structural Statistics

Statistic	SA	SA,
RMSDs from experimental distance restraints (angstroms)[a]		
All (1084)	0.028 ± 0.003	0.026
Sequential (327)	0.048 ± 0.006	0.027
Short range $(2 \leqslant |i - j| \leqslant 4)$ (57)	0.021 ± 0.011	0.012
Long range $(|i - j| > 4)$ (316)	0.033 ± 0.005	0.030
H bond (66)[b]	0.046 ± 0.010	0.049
RMS deviations from experimental dihedral restraints (degrees)	0.284 ± 0.070	0.163
Deviations from idealized covalent geometry		
Bonds (angstroms)	0.008 ± 0.0004	0.008
Angles (degrees)	2.094 ± 0.008	2.448
Impropers (degrees)	1.034 ± 0.014	1.035
Energies (kJ/mol)		
F_{NOE}[c]	121.7 ± 29.1	100.3
F_{cdih}	2.3 ± 1.2	0.5
F_{bond}	175.9 ± 9.6	181.6
F_{angle}	7161.2 ± 52.3	9794.4
$F_{improper}$	396.9 ± 10.5	396.8
F_{repel}[d]	142.9 ± 17.2	155.7
F_{LJ}[e]	-877.6 ± 46.6	763.0

Atomic RMS differences (angstroms)

	Backbone atoms	All heavy atoms
Residues 1–94	1.70 ± 0.52	2.03 ± 0.4
Residues 8–94	1.33 ± 0.28	1.71 ± 0.25
Residues 8–40, 46–77, 86–94 (without N-terminus and flexible loops)	0.83 ± 0.26	1.22 ± 0.16
Residues 9–14, 17–23, 31–38, 46–51, 56–60, 66–76, 88–94 (β strands only)	0.52 ± 0.11	0.98 ± 0.10

Structure notation: SA refers to the 36 refined simulated annealing structures; SA, refers to the energy-minimized average structure. This structure was obtained by averaging the coordinates of the final structures best-fitted to each other over the backbone atoms of residues 8–94, with the resulting coordinates minimized with the experimental restraints.

[a] None of the final structures exhibit distance restraint violations >0.4Å or dihedral angle restraint violations >3°: the RMSDs are calculated relative to the mean structure.

[b] Each hydrogen bond is characterized by two distance restraints: $d_{NH-O} < 2.3$ Å, $d_{N-O} < 3.3$ Å. Hydrogen bond restraints were only included when they could be shown to be part of a regular secondary structure element.

[c] The final values of the square-well NOE and torsion angle potentials are calculated with force constants of 210 kJmol per Å^{-2} and 840 kJ mol per rad^{-2}, respectively.

[d] The quadratic van der Waals term (F_{repel}) is calculated with a force constant of 17 kJmol per Å^{-4} with the van der Waals radii set to 0.8 times the standard values used in the CHARMM empirical energy function (Brooks et al., 1983).

[e] F_{LJ} is calculated using the full CHARMM empirical energy function. This term is not included in the target function during refinement, so it provides an indication of the quality of nonbonded interactions in the structures.

types often possess highly conserved Gly residues, which are necessary for the formation of certain types of tight β turns (Wilmot and Thornton, 1990). The type III module has only one tight β turn that does not require a Gly residue. A third instance where residues may be conserved for structural purposes is the correct formation of interfaces between sheets or between modules; Pro-25, the loop between the sheets that joins strands E and F, and the proline residue near the N-terminus of the module may belong to this category. The connection between strands E and F is a conserved five-membered loop in all the type III modules in fibronectin; the first and last residues of the loop show a marked Gly preference. Similarly, the turn between strands A and B is of consistent length. The remaining loops of the module are all highly variable in length, the insertion of the RGDS sequence in the F–G loop being particularly striking.

Dynamic Behavior

Figure 4 shows the NOE distribution (Figure 4A) and the root mean square deviation (RMSD) per residue on super-position of the backbone atoms (Figure 4B). The lack of long-range NOEs observed for residues in the loops between strands C and C' (residues 39–45) and strands F and G (residues 77–87) leads to high RMSDs in the calculated structures, suggesting that these loops are conformationally labile. The results of a heteronuclear $^{15}N-^{1}H$ NOE experiment are shown in Figure 4C. The size of the NOE observed reflects the dynamic behavior of the module (Kay et al., 1989); a smaller NOE is observed for more mobile parts of the molecule. The heteronuclear NOEs for residues in the C–C' and F–G loops are significantly smaller than those of the majority of the molecule, which indicates considerable conformational flexibility in these regions. The heteronuclear NOEs for residues Gly-79, Asp-80, Ser-
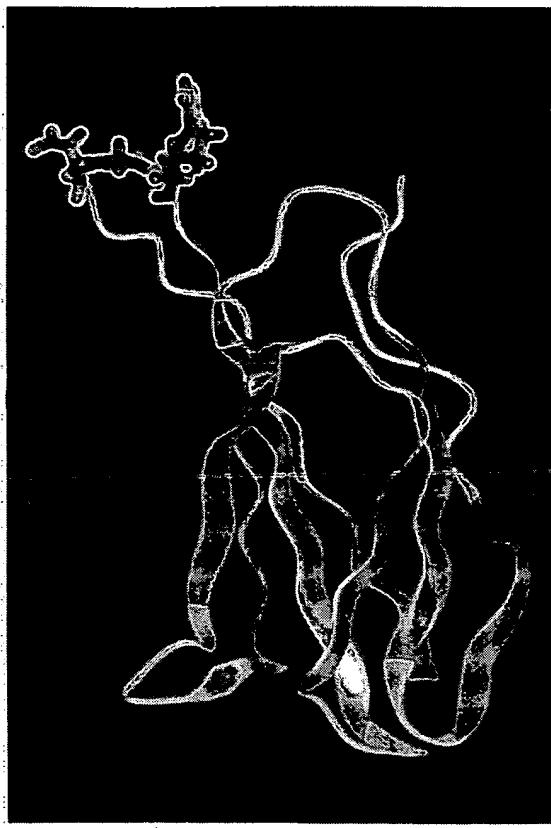
Figure 2. Ribbon Representation of the Module, Highlighting Secondary Structure Elements and the RGD Motif

The backbone of the energy-minimized average structure is shown with β strands colored red and the loops colored blue. The side chains of Arg-78, Gly-79, and Asp-80 are colored yellow and are clearly seen to be at the apex of the loop, protruding from the body of the module. As described fully later in this paper, the RGD tripeptide does not appear to have a fixed conformation in solution; thus, the stylized side chains are included for clarity and ease of interpretation only.

81, Lys-83, and Ser-84 are all of similar magnitude, which suggests that they are undergoing similar amplitudes of motion; this may imply that the loop undergoes some sort of conformational equilibrium, possibly a hinge motion, rather than being completely disordered. The flexibility of the two loops is likely to be the cause of the lack of homonuclear NOEs and higher RMSDs for these residues. Although the lack of NOE restraints could have resulted from difficulties in interpretation of the spectra as a consequence of overlapped peaks, the heteronuclear experiment gives an independent measure of the dynamic properties of the molecule, showing that the loops are genuinely flexible and substantiating the conclusions drawn from the distribution of NOE distance restraints alone. The β strands are much less flexible and appear to provide a rigid framework upon which functional, flexible loops are built. To obtain a more detailed picture of the dynamic behavior of the type III module, heteronuclear NMR experiments are underway to extract precise order parameters and also to probe the interaction of the module with a short, synthetic integrin peptide.

Comparison with Other Known Structures

An interesting feature of the structures present in the Protein Data Bank is that in a nonredundant set of 254 structures, only 83 had unique folds (Pascarella and Argos, 1992); in addition, structures with no detectable sequence homology were found to possess the same fold. The sequence and structure of the type III module have been compared with other proteins of known structure to ascertain whether this represents a new fold or another member of an established family. The topology is similar to that of immunoglobulin C domains (Williams and Barclay, 1988); however, strand C′ is hydrogen bonded to strand C rather than to strand E (Figure 5a). This alternative strand arrangement has also been observed in domain 2 (D2) of the T cell glycoprotein CD4 (Wang et al., 1990; Ryu et al., 1990), the D2 of the bacterial chaperone protein PapD

Figure 3. Sequence Alignment of the Type III Modules of Human Fibronectin

The alignment of the 16 type III modules of human fibronectin described by Kornblihtt et al. (1985) is shown. F12 corresponds to the ED-A sequence, which is not always present in the protein as a consequence of alternative splicing of the mRNA. The highly conserved residues are indicated by asterisks.
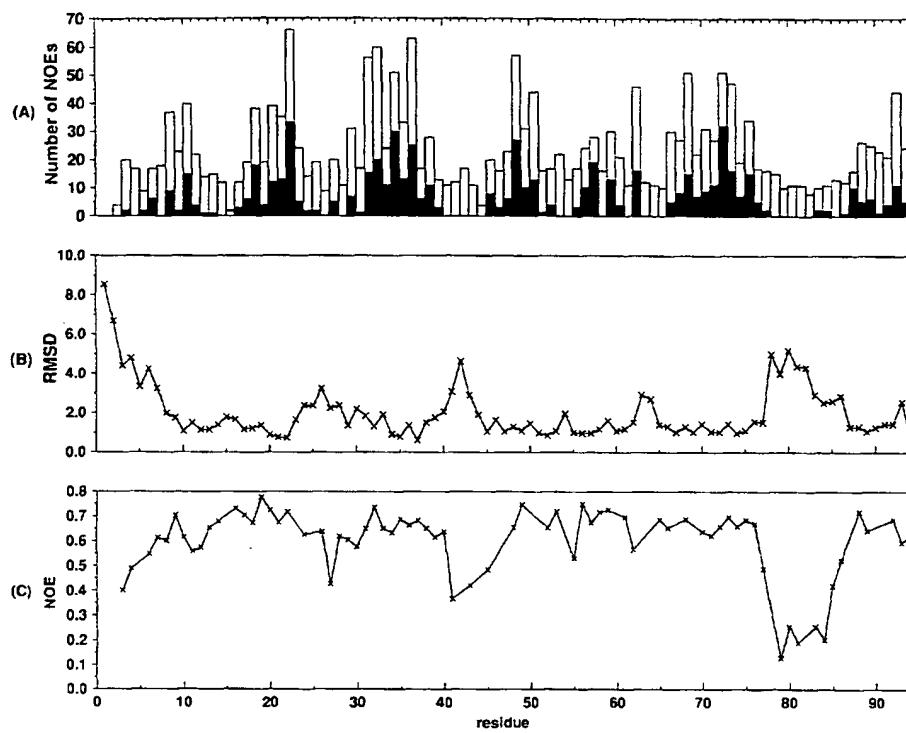
Figure 4. The Dynamic Behavior of the Module

(A) The distribution of NOEs is shown, with all NOEs and long-range NOEs |i–j| > 4 indicated by the open and closed bars, respectively. There is a notable lack of long-range NOEs from the residues in the loops between strands C and C' (Gly-41–Val-45) and strands F and G (Arg-78–Lys-86).

(B) Plot of the average RMSD (in angstroms) of the backbone atoms as a function of residue number, after the superposition of the backbone atoms of residues 8–94 of the 36 final structures.

(C) Plot of the backbone amide ^{15}N–^{1}H heteronuclear NOEs as a function of residue number. A number of points are missing as a consequence of spectral overlap. All Pro residues and Val-1 are not observed.
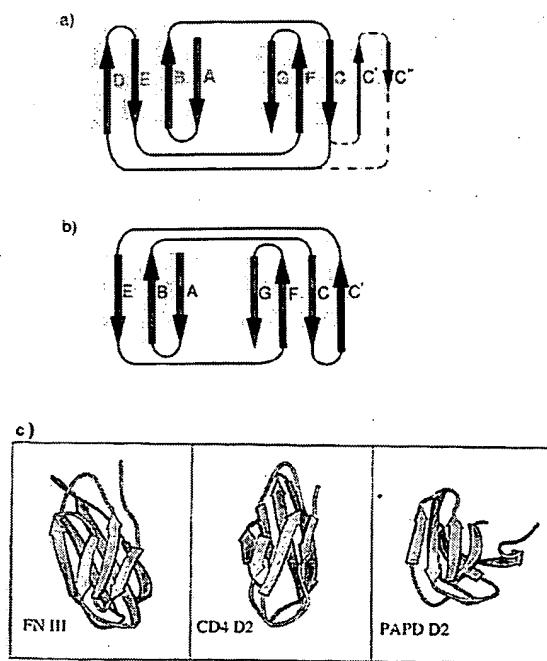

Figure 5. Folds of Similar Topology to the Type III Module

(Holmgren and Bränden, 1989), and the two domains that constitute the extracellular region of the human growth hormone receptor (hGHR) (de Vos et al., 1992).

There is no significant sequence similarity between the fibronectin type III module described herein and the domains of similar topology found in PapD and CD4. The alignment scores for these modules, obtained using the program ALIGN (Dayhoff et al., 1983), were 0.61 and −0.16 standard deviation units respectively; alignment scores of 5 standard deviation units or higher are usually taken to indicate clear sequence similarity (Barton, 1990). In spite of the lack of sequence similarity, there is clear topological similarity. The coordinates of PapD and CD4 are available and MOLSCRIPT diagrams of the type III module, CD4 D2, and PapD D2 are shown in Figure 5b.

(a) Illustration of the topology of immunoglobulin domains, with the seven strands of C domains shown with solid lines and the additional C strands of V domains shown with dashed lines.

(b) Illustration of the topology of the fibronectin type III module, which is also observed in CD4 D2, PapD D2, and the hGHR domains.

(c) MOLSCRIPT diagrams (Kraulis, 1991) of the type III structures of fibronectin, CD4 D2, and PapD D2, demonstrating the clear topological similarity, but the markedly different global fold.

Upon alignment with 38 core Cα atoms of the type III module, RMSD values of 4.4 and 2.3, respectively, were obtained. The poor overlay, both quantitatively described above and qualitatively shown in Figure 5b, may suggest that the modules described are related by convergent rather than divergent evolution. Bazan detected the structural similarity between cytokine modules and fibronectin type III (Bazan, 1990); however, the coordinates of hGHR are not available, and, thus, a detailed comparison of the structures cannot be carried out.

Two cytokine receptor modules (hGHR 1 and hGHR 2) of the hGHR provide an example where structural similarity to the fibronectin type III module has been predicted, by the analysis of patterns of hypdrophobic and hydrophilic residues (Bazan, 1990). The structures of these modules have recently been determined in the X-ray crystallography structure of hGH bound to a receptor dimer (de Vos et al., 1992); it is clear that they have similar topology to the module structure determined here.

Implications for Integrin Binding

The heteronuclear NOE experiment, described in this paper, provides direct evidence that the loops bearing RGD sequences may be flexible, rather than existing in a specific conformation. The RGD motifs of potent integrin inhibitors of known structure also lie at the apex of conformationally flexible loops (Saudek et al., 1991; Adler et al., 1991); however, in these cases, dynamic properties were only inferred from the lack of identified restraints. The presence of RGD sequences at the apex of solvent-exposed, flexible loops suggests that they may be responsible for fast recognition and fitting to the receptor.

In recent years it has become apparent that the RGD sequence alone does not account for the full cell adhesive properties of fibronectin and that additional synergistic regions are required for full activity (Yamada, 1991). Using internal deletion and 5' terminal deletion mutants, Aota et al. (1991) located two distinct peptide regions, in the eighth and ninth type III modules, that contributed to the adhesive capacity of fibronectin. Removal of the region at the center of the ninth type III module resulted in the greatest loss of activity. By homology, this central region would correspond to strands C and C' of the module and would include the loop between these strands. The alignment of the type III modules of fibronectin suggests that the C–C' loop of the ninth module is the same length as the module described in this paper. It is, therefore, likely to have similar dynamic properties and show considerable flexibility. The location of the loop between strands C and C' of the ninth type III module as the synergistic site is not inconsistent with the results of monoclonal antibody studies (Aota et al., 1991; Nagai et al., 1991). This loop may lie at the domain–domain interface between the ninth and tenth type III modules and is a viable candidate for interaction with either the RGD loop or the integrin, depending on the exact nature of the module–module orientation. Thus, it is possible that the C–C' loop is responsible for the synergy observed.

The recent elegant work of de Vos et al. (1992) on hGH binding to a receptor dimer gives possible insight into protein–protein interactions that involve more than one type III module. In the complex, one hormone molecule interacts with two cytokine receptor molecules. The four cytokine receptor modules all contribute residues that participate in hormone binding. The receptor binding interface consists of loops A–B and E–F of hGHR 1 and loops F–G and B–C of hGHR 2. Similarly, both cytokine receptor modules of the interleukin 3–binding protein have been shown to be involved in interleukin 3 binding (Wang et al., 1992). Thus, the participation of two adjacent type III modules in the interaction between fibronectin and its integrin receptors may be similar to the behavior observed in the structurally related cytokine receptor modules.

The binding of fibronectin to its integrin receptors is further complicated by the fact that integrin affinity and specificity can be modulated as a consequence of events within cells (Hynes, 1992). Clearly, the interaction between fibronectin and its receptors can be finely tuned to encompass the wide range of integrin interactions among different cells. The structure presented in this paper gives insight into the way a functional loop can be built onto a structural framework and, by virtue of its flexibility, be able to perform a wide range of functions.

Experimental Procedures

The tenth type III module (corresponding to residues 1416–1509 of human fibronectin [Kornblihtt et al., 1985] and referred to as residues 1–94 in this paper) was expressed using a yeast secretion system based on the α factor leader sequence and purified to homogeneity as described in Baron et al. (1992). In brief, the purification involves adsorption onto C18 reverse-phase beads (supplied by high performance liquid chromatography technology), elution with 60% acetonitrile, 0.1% trifluoroacetic acid, and lyophilization. The protein is then redissolved in water and further purified by a combination of reverse-phase and cation-exchange high performance liquid chromatography.

For the preparation of NMR samples, the protein was dissolved in either D_2O or 90% H_2O, 10% D_2O to a concentration of 3 mM. NMR spectra were recorded at pH 3.9 and at temperatures of 20°C, 39°C, and 47°C. Proton–proton distances were determined from nuclear Overhauser enhancements measured in NOESY experiments. Mixing times of 50 and 160 ms for D_2O NOESY spectra and 75 and 200 ms for H_2O NOESY were used to assess proton–proton distances. Upper limits for distance restraints were categorized according to the estimated intensity of the NOE cross peak in the spectrum: upper limits of 2.5 Å, 3.3 Å, and 5.0 Å for strong, medium, and weak peaks, respectively, were used. Appropriate corrections were added when the NOEs involved degenerate proton resonances (Wüthrich, 1986). The lower limits were set explicitly to 0.0 Å, which is more appropriate when using simulated annealing protocols for structure calculation (Hommel et al., 1992).

Slowly exchanging amide NH groups were identified from a 7 hr homonuclear Hartmann–Hahn experiment (Braunschweiler and Ernst, 1983; Davis and Bax, 1985) recorded immediately after dissolving the protein in D_2O. Hydrogen bonds that were present in β sheet regions were included in the structure calculations when a hydrogen bond acceptor could be unambiguously assigned, using NOE data characteristic of regular secondary structure. Each hydrogen bond was incorporated as a pair of distance restraints, again, with only upper limits defined ($d_{NH-O} < 2.3$ Å, $d_{N-O} < 3.3$ Å).

$^3J_{HN-\alpha CH}$ coupling constants were determined using a ^1H detected heteronuclear ^1H–^{15}N multiple quantum coherence spectrum (HMQC-J) by fitting F1 traces to a theoretical lineshape (Kay and Bax, 1990). $^3J_{\alpha CH-\beta CH}$ coupling constants were determined using the passive coupling of αCH–βCH cross peaks in a P. E. COSY spectrum (Mueller, 1987; Marion and Bax, 1988). STEREOSEARCH (Nilges et al., 1990), a method involving the search of a systematic data base, was used to obtain φ, ψ, and χ_1 dihedral angle restraints from the measured cou-

pling constants and short intraresidue ($d_{\text{NH+}\alpha\text{CH}}$, $d_{\alpha\text{CH+}\beta\text{CH}}$) and sequential ($d_{\alpha\text{CH+NH}}$, $d_{\beta\text{CH+NH}}$) distances. For ϕ, ψ, and χ_1, the minimum ranges allowed were $\pm 30°$, $\pm 50°$, and $\pm 20°$, respectively (Kraulis et al., 1989).

Structure calculations were carried out using the program XPLOR (Brünger, 1988). In brief, initial structures are calculated using a dynamical simulated annealing method, starting from a structure with random backbone dihedral angles and extended side chains (Nilges et al., 1991). These initial structures were then subjected to two rounds of refinement using a simulated annealing protocol (Downing et al., 1992; Hommel et al., 1992). An energy-minimized average structure was determined by calculating the mean coordinate positions of the 36 final structures best-fitted on the backbone atoms of residues 8–94, followed by restrained minimization of the same target function as that used in the final stages of the structure calculation.

The multiple sequence alignment of the fibronectin type III modules was generated using the Alignment of Multiple Protein Sequences program of Barton and Sternberg (1987). A bias of 8 was added to each term of the mutation data matrix (Dayhoff et al., 1983) and a gap penalty of 6 was used. One hundred random runs were performed to establish mean random scores. The sequences were ordered and aligned using a "Tree" method. Pairwise comparisons of the tenth type III module with the second domains of CD4 and PapD were carried out using the program ALIGN (Dayhoff et al., 1983) with a bias of 6, a gap penalty of 6, and 100 random runs. The alignment scores are the distance in standard deviation units of the score for the pairwise comparison from the mean random score for that pair.

References

Adler, M. A., Lazarus, R. A., Dennis, M. S., amd Wagner, G. (1991). Solution structure of kistrin, a potent platelet aggregation inhibitor and GPIIb-IIIa antagonist. Science 253, 445–448.

Aota, S.-i., Nagai, T., and Yamada, K. M. (1991). Characterization of regions of fibronectin besides the arginine-glycine-aspartic acid sequence required for adhesive function of the cell-binding domain using site directed mutagenesis. J. Biol. Chem. 266, 15938–15943.

Baron, M., Norman, D., Willis, A., and Campbell, I. D. (1990). Structure of the fibronectin type 1 module. Nature 345, 642–646.

Baron, M., Norman, D. G., and Campbell, I. D. (1991). Protein modules. Trends Biochem. Sci. 16, 13–17.

Baron, M., Main, A. L., Driscoll, P. C., Mardon, H. J., Boyd, J., and Campbell, I. D. (1992). ¹H NMR assignment and secondary structure of the cell adhesion type III module of fibronectin. Biochemistry 31, 2068–2073.

Barton, G. J. (1990). Protein multiple sequence alignment and flexible pattern matching. Meth. Enzymol. 183, 403–428.

Barton, G. J., and Sternberg, M. J. E. (1987). A strategy for the rapid multiple alignment of protein structures: confidence levels from tertiary structure comparisons. J. Mol. Biol. 198, 327–337.

Bazan, J. F. (1990). Structural design and molecular evolution of a cytokine receptor superfamily. Proc. Natl. Acad. Sci. USA 87, 6934–6938.

Braunschweiler, L. R., and Ernst, R. R. (1983). Coherence transfer by isotropic mixing: application to protein correlation spectroscopy. J. Magn. Reson. 53, 521–528.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983). CHARMm: a program for macromolecular energy, minimization, and dynamic calculations. J. Comput. Chem. 4, 187–217.

Brünger, A. T. (1988). XPLOR Manual (New Haven: Yale University).

Constantine, K. L., Madrid, M., Banyai, L., Trexler, M., Patthy, L., and Llinas, M. (1992). Refined solution structure and ligand-binding properties of PDC-109 domain b: a collagen-binding type II domain. Biochemistry 30, 1663–1672.

Davis, D. G., and Bax, A. (1985). Assignment of complex ¹H n.m.r. spectra via two-dimensional homonuclear Hartmann-Hahn spectroscopy. J. Am. Chem. Soc. 107, 2820–2821.

Dayhoff, M. O., Barker, W. C., and Hunt, L. T. (1983). Establishing homologies in protein sequences. Meth. Enzymol. 91, 524–545.

de Vos, A. M., Ultsch, M., and Kossiakoff, A. A. (1992). Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. Science 255, 306–312.

Downing, A. K., Driscoll, P. C., Harvey, T. S., Dudgeon, T. J., Smith, B. O., Baron, M., and Campbell, I. D. (1992). The solution structure of the fibrin binding finger domain of tissue-type plasminogen activator determined by ¹H nuclear magnetic resonance. J. Mol. Biol. 225, 821–833.

D'Souza, S. E., Ginsberg, M. H., and Plow, E. F. (1991). Arginyl-glycyl-aspartic acid (RGD): a cell adhesion motif. Trends Biochem. Sci. 16, 246–250.

Güntert, P., Braun, W., and Wüthrich, K. (1991). Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. J. Mol. Biol. 217, 517–530.

Holmgren, A., and Bränden, C.-I. (1989). Crystal structure of chaperone protein PapD reveals an immunoglobulin fold. Nature 342, 248–251.

Hommel, U., Harvey, T. S., Driscoll, P. C., and Campbell, I. D. (1992). Human epidermal growth factor: high resolution solution structure and comparison with human transforming growth factor α. J. Mol. Biol., in press.

Hynes, R. O. (1990). Fibronectins (New York: Springer Verlag).

Hynes, R. O. (1992). Integrins: versatility, modulation, and signaling in cell adhesion. Cell 69, 11–25.

Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22, 2577–2637.

Kay, L. E., and Bax, A. (1990). New methods for the measurement of NH-CαH coupling constants in ¹⁵N-labeled proteins. J. Magn. Reson. 86, 110–126.

Kay, L. E., Torchia, D. A., and Bax, A. (1989). Backbone dynamics of proteins as studied by ¹⁵N inverse detected heteronuclear NMR spectroscopy: application to staphlyococcal nuclease. Biochemistry 28, 8972–8979.

Kornblihtt, A. R., Umezawa, K., Vibe-Pederson, K., and Baralle, F. E. (1985). Primary structure of human fibronectin: differential splicing may generate at least 10 polypeptides from a single gene. EMBO J. 4, 1755–1759.

Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Cryst. 24, 946–950.

Kraulis, P. J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J., and Gronenborn, A. M. (1989). Determination of the three-dimensional structure of the C-terminal domain of cellobiohydrolase I from Trichoderma reesei. A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. Biochemistry 28, 7241–7257.

Labeit, S., Barlow, D. P., Gautel, M., Gibson, T., Holt, J., Hsieh, C.-L., Francke, U., Leonard, K., Wardale, J., Whiting, A., and Trinick, J. (1990). A regular pattern of two types of 100-residue motif in the sequence of titin. Nature 345, 273–276.

Marion, D., and Bax, A. (1988). P. COSY, a sensitive alternative for double-quantum-filtered COSY. J. Magn. Reson. 80, 528–533.

Mayford, M., Barzilai, A., Keller, F., Schacher, S., and Kandel, E. R.

(1992). Modulation of an NCAM-related adhesion molecule with long-term synaptic plasticity in *Aplysia*. Science *256*, 638–644.

Mosher, D. F. (1989). Fibronectin (New York: Academic Press).

Mueller, L. (1987). P. E. COSY a simple alternative to E. COSY. J. Magn. Reson. *72*, 191–196.

Nagai, T., Yamakawa, N., Aota, S.-i., Yamada, S. S., Akiyama, S. K., Olden, K., and Yamada, K. M. (1991). Monoclonal antibody characterization of two distant sites required for function of the central cell-binding domain of fibronectin in cell adhesion, cell migration and matrix assembly. J. Cell Biol. *114*, 1295–1305.

Nilges, M., Clore, G. M., and Gronenborn, A. M. (1990). 'H-NMR stereospecific assignments by conformational data-base searches. Biopolymers *29*, 813–822.

Nilges, M., Kuszewski, J., and Brünger, A. T. (1991). Sampling properties of simulated annealing and distance geometry. In Computational Aspects of the Study of Biological Macromolecules, J. C. Hoch, ed. (New York: Plenum Press), pp. 451–455.

Obara, M., Kang, M. S., and Yamada, K. M. (1988). Site-directed mutagenesis of the cell-binding domain of human fibronectin: separable, synergistic sites mediate adhesive function. Cell *53*, 649–657.

Pascarella, S., and Argos, P. (1992). A data bank merging related protein structures and sequences. Prot. Eng. *5*, 121–137.

Patthy, L. (1991). Modular exchange principles in proteins. Curr. Opin. Struct. Biol. *1*, 351–361.

Pierschbacher, M. D., and Ruoslahti, E. (1984). Cell attachment activity of fibronectin can be duplicated by small synthetic fragments of the molecule. Nature *309*, 30–33.

Richardson, J. S., Getzoff, E. D., and Richardson, D. G. (1978). The β bulge: a common small unit of nonrepetitive protein structure. Proc. Natl. Acad. Sci. USA *75*, 2574–2578.

Ruoslahti, E. (1988). Fibronectin and its receptors. Annu. Rev. Biochem. *57*, 375–413.

Ruoslahti, E., and Pierschbacher, M. D. (1987). New perspectives in cell adhesion: RGD and integrins. Science *238*, 491–497.

Ryu, S.-E., Kwong, P. D., Truneh, A., Porter, T. G., Arthos, J., Rosenberg, M., Dai, X., Xuong, N.-h., Axel, R., Sweet, R. W., and Hendrickson, W. A. (1990). Crystal structure of an HIV-binding recombinant fragment of human CD4. Nature *348*, 419–426.

Saudek, V., Atkinson, R. A., and Pelton, J. T. (1991). Three-dimensional structure of echistatin, the smallest active RGD protein. Biochemistry *30*, 7369–7372.

Sibanda, B. L., Blundell, T. L., and Thornton, J. M. (1989). Conformation of β-hairpins in protein structures: a systematic classification with applications to modelling by homology, electron density fitting and protein engineering. J. Mol. Biol. *206*, 759–777.

Wang, J., Yan, Y., Garrett, T. P. J., Liu, J., Rodgers, D. W., Garlick, R. L., Tarr, G. E., Husain, Y., Reinherz, E. L., and Harrison, S. C. (1990). Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains. Nature *348*, 411–418.

Wang, H.-M., Ogorochi, T., Arai, K.-i., and Miyajima, A. (1992). Structure of mouse interleukin 3 (IL-3) binding protein (Al2CA): amino acid residues critical for IL-3 binding. J. Biol. Chem. *267*, 979–983.

Williams, A. F., and Barclay, A. N. (1988). The immunoglobulin superfamily—domains for cell surface recognition. Annu. Rev. Immunol. *6*, 381–405.

Wilmot, C. M., and Thornton, J. M. (1990). β-Turns and their distortions: a proposed new nomenclature. Prot. Eng. *3*, 479–493.

Wüthrich, K. (1986). NMR of Proteins and Nucleic Acids (New York: John Wiley & Sons, Inc.).

Yamada, K. M. (1991). Adhesive recognition sequences. J. Biol. Chem. *266*, 12809–12812.

(51) International Patent Classification [6] : C12N 15/12, C07K 14/78, C12N 15/70, 1/21	A2	(11) International Publication Number: **WO 98/56915**
		(43) International Publication Date: 17 December 1998 (17.12.98)

(54) Title: ARTIFICIAL ANTIBODY POLYPEPTIDES

```
      NdeI                            PstI                    EcoRI
        1         11        21        31        41
      mq VSDVPRDLEV VAATPTSLLI SWDAPAVTVR YYRITYGETG GNSPVQEFTV
              A         B                    C           D

                   SalI                 SacI      XhoI
      51        61        71        81        91
        PGSKSTATIS GLKPGVDYTI TVYAVTGRGD SPASSKPISI NYRT
              E          F                    G
```

(57) Abstract

A fibronectin type III (Fn3) polypeptide monobody, a nucleic acid molecule encoding said monobody, and a variegated nucleic acid library encoding said monobody, are provided by the invention. Also provided are methods of preparing a Fn3 polypeptide monobody, and kits to perform said methods. Further provided is a method of identifying the amino acid sequence of a polypeptide molecule capable of binding to a specific binding partner (SBP) so as to form a polypeptide:SSP complex, and a method of identifying the amino acid sequence of a polypeptide molecule capable of catalyzing a chemical reaction with a catalyzed rate constant, k_{cat}, and an uncatalyzed rate constant, k_{uncat}, such that the ratio of k_{cat}/k_{uncat} is greater than 10.

ARTIFICIAL ANTIBODY POLYPEPTIDES

FIELD OF THE INVENTION

The present invention relates generally to the field of the production and selection of binding and catalytic polypeptides by the methods of molecular biology, using both combinatorial chemistry and recombinant DNA. The invention specifically relates to the generation of both nucleic acid and polypeptide libraries derived therefrom encoding the molecular scaffolding of Fibronectin Type III (Fn3) modified in one or more of its loop regions. The invention also relates to the "artificial mini-antibodies" or "monobodies," i.e., the polypeptides comprising an Fn3 scaffold onto which loop regions capable of binding to a variety of different molecular structures (such as antibody binding sites) have been grafted.

BACKGROUND OF THE INVENTION

Antibody structure

A standard antibody (Ab) is a tetrameric structure consisting of two identical immunoglobulin (Ig) heavy chains and two identical light chains. The heavy and light chains of an Ab consist of different domains. Each light chain has one variable domain (VL) and one constant domain (CL), while each heavy chain has one variable domain (VH) and three or four constant domains (CH) (Alzari et al., 1988). Each domain, consisting of ~110 amino acid residues, is folded into a characteristic β-sandwich structure formed from two β-sheets packed against each other, the immunoglobulin fold. The VH and VL domains each have three complementarity determining regions (CDR1-3) that are loops, or turns, connecting β-strands at one end of the domains (Fig. 1: A, C). The variable regions of both the light and heavy chains generally contribute to antigen specificity, although the contribution of the individual chains to specificity is not always equal. Antibody molecules have evolved to bind to a large number of molecules by using six randomized loops (CDRs). However, the size of the antibodies and the complexity of six loops represents a major design hurdle if the end result is to be a relatively small peptide ligand.

2

Antibody substructures

Functional substructures of Abs can be prepared by proteolysis and by recombinant methods. They include the Fab fragment, which comprises the VH-CH1 domains of the heavy chain and the VL-CL1 domains of the light chain

5   joined by a single interchain disulfide bond, and the Fv fragment, which comprises only the VH and VL domains. In some cases, a single VH domain retains significant affinity (Ward et al., 1989). It has also been shown that a certain monomeric κ light chain will specifically bind to its cognate antigen. (L. Masat et al., 1994). Separated light or heavy chains have sometimes been found

10   to retain some antigen-binding activity (Ward et al., 1989). These antibody fragments are not suitable for structural analysis using NMR spectroscopy due to their size, low solubility or low conformational stability.

Another functional substructure is a single chain Fv (scFv), comprised of the variable regions of the immunoglobulin heavy and light chain, covalently

15   connected by a peptide linker (S-z Hu et al., 1996). These small (M_r 25,000) proteins generally retain specificity and affinity for antigen in a single polypeptide and can provide a convenient building block for larger, antigen-specific molecules. Several groups have reported biodistribution studies in xenografted athymic mice using scFv reactive against a variety of tumor

20   antigens, in which specific tumor localization has been observed. However, the short persistence of scFvs in the circulation limits the exposure of tumor cells to the scFvs, placing limits on the level of uptake. As a result, tumor uptake by scFvs in animal studies has generally been only 1-5%ID/g as opposed to intact antibodies that can localize in tumors ad 30-40 %ID/g and have reached levels as

25   high as 60-70 %ID/g.

A small protein scaffold called a "minibody" was designed using a part of the Ig VH domain as the template (Pessi et al., 1993). Minibodies with high affinity (dissociation constant (K_d) ~10^{-7} M) to interleukin-6 were identified by randomizing loops corresponding to CDR1 and CDR2 of VH and then selecting

30   mutants using the phage display method (Martin et al., 1994). These experiments demonstrated that the essence of the Ab function could be

transferred to a smaller system. However, the minibody had inherited the limited solubility of the VH domain (Bianchi et al., 1994).

It has been reported that camels (*Camelus dromedarius*) often lack variable light chain domains when IgG-like material from their serum is

5   analyzed, suggesting that sufficient antibody specificity and affinity can be derived form VH domains (three CDR loops) alone. Davies and Riechmann recently demonstrated that "camelized" VH domains with high affinity ($K_d \sim 10^-$

^{7} M) and high specificity can be generated by randomizing only the CDR3. To improve the solubility and suppress nonspecific binding, three mutations were

10   introduced to the framework region (Davies & Riechmann, 1995). It has not been definitively shown, however, that camelization can be used, in general, to improve the solubility and stability of VHs.

An alternative to the "minibody" is the "diabody." Diabodies are small bivalent and bispecific antibody fragments, i.e., they have two antigen-binding

15   sites. The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) on the same polypeptide chain (V_H-V_L). Diabodies are similar in size to an Fab fragment. By using a linker that is too short to allow pairing between the two domains on the same chain, the domains are forced to pair with the complementary domains of another chain and create

20   two antigen-binding sites. These dimeric antibody fragments, or "diabodies," are bivalent and bispecific. P. Holliger et al., PNAS 90:6444-6448 (1993).

Since the development of the monoclonal antibody technology, a large number of 3D structures of Ab fragments in the complexed and/or free states have been solved by X-ray crystallography (Webster et al., 1994; Wilson &

25   Stanfield, 1994). Analysis of Ab structures has revealed that five out of the six CDRs have limited numbers of peptide backbone conformations, thereby permitting one to predict the backbone conformation of CDRs using the so-called canonical structures (Lesk & Tramontano, 1992; Rees et al., 1994). The analysis also has revealed that the CDR3 of the VH domain (VH-CDR3) usually

30   has the largest contact surface and that its conformation is too diverse for canonical structures to be defined; VH-CDR3 is also known to have a large

4

variation in length (Wu et al., 1993). Therefore, the structures of crucial regions of the Ab-antigen interface still need to be experimentally determined.

Comparison of crystal structures between the free and complexed states has revealed several types of conformational rearrangements. They include side-

5   chain rearrangements, segmental movements, large rearrangements of VH-CDR3 and changes in the relative position of the VH and VL domains (Wilson & Stanfield, 1993). In the free state, CDRs, in particular those which undergo large conformational changes upon binding, are expected to be flexible. Since X-ray crystallography is not suited for characterizing flexible parts of molecules,

10   structural studies in the solution state have not been possible to provide dynamic pictures of the conformation of antigen-binding sites.

Mimicking the antibody-binding site

CDR peptides and organic CDR mimetics have been made (Dougall et al., 1994). CDR peptides are short, typically cyclic, peptides which correspond

15   to the amino acid sequences of CDR loops of antibodies. CDR loops are responsible for antibody-antigen interactions. Organic CDR mimetics are peptides corresponding to CDR loops which are attached to a scaffold, e.g., a small organic compound.

CDR peptides and organic CDR mimetics have been shown to retain

20   some binding affinity (Smyth & von Itzstein, 1994). However, as expected, they are too small and too flexible to maintain full affinity and specificity. Mouse CDRs have been grafted onto the human Ig framework without the loss of affinity (Jones et al., 1986; Riechmann et al., 1988), though this "humanization" does not solve the above-mentioned problems specific to solution studies.

25   ## Mimicking natural selection processes of Abs

In the immune system, specific Abs are selected and amplified from a large library (affinity maturation). The processes can be reproduced *in vitro* using combinatorial library technologies. The successful display of Ab fragments on the surface of bacteriophage has made it possible to generate and

30   screen a vast number of CDR mutations (McCafferty et al., 1990; Barbas et al., 1991; Winter et al., 1994). An increasing number of Fabs and Fvs (and their

5

derivatives) is produced by this technique, providing a rich source for structural
studies. The combinatorial technique can be combined with Ab mimics.

A number of protein domains that could potentially serve as protein
scaffolds have been expressed as fusions with phage capsid proteins. Review in
5   Clackson & Wells, Trends Biotechnol. 12:173-184 (1994). Indeed, several of
these protein domains have already been used as scaffolds for displaying random
peptide sequences, including bovine pancreatic trypsin inhibitor (Roberts et al.,
PNAS 89:2429-2433 (1992)), human growth hormone (Lowman et al.,
Biochemistry 30:10832-10838 (1991)), Venturini et al., Protein Peptide Letters
10   1:70-75 (1994)), and the IgG binding domain of *Streptococcus* (O'Neil et al.,
Techniques in Protein Chemistry V (Crabb, L,. ed.) pp. 517-524, Academic
Press, San Diego (1994)). These scaffolds have displayed a single randomized
loop or region.

Researchers have used the small 74 amino acid α-amylase inhibitor
15   Tendamistat as a presentation scaffold on the filamentous phage M13
(McConnell and Hoess, 1995). Tendamistat is a β-sheet protein from
Streptomyces tendae. It has a number of features that make it an attractive
scaffold for peptides, including its small size, stability, and the availability of
high resolution NMR and X-ray structural data. Tendamistat's overall topology
20   is similar to that of an immunoglobulin domain, with two β-sheets connected by
a series of loops. In contrast to immunoglobulin domains, the β-sheets of
Tendamistat are held together with two rather than one disulfide bond,
accounting for the considerable stability of the protein. By analogy with the
CDR loops found in immunoglobulins, the loops the Tendamistat may serve a
25   similar function and can be easily randomized by in vitro mutagenesis.

Tendamistat, however, is derived from *Streptomyces tendae*. Thus,
while Tendamistat may be antigenic in humans, its small size may reduce or
inhibit its antigenicity. Also, Tendamistat's stability is uncertain. Further, the
stability that is reported for Tendamistat is attributed to the presence of two
30   disulfide bonds. Disulfide bonds, however, are a significant disadvantage to
such molecules in that they can be broken under reducing conditions and must be

6

properly formed in order to have a useful protein structure. Further, the size of
the loops in Tendamistat are relatively small, thus limiting the size of the inserts
that can be accommodated in the scaffold. Moreover, it is well known that
forming correct disulfide bonds in newly synthesized peptides is not

5   straightforward. When a protein is expressed in the cytoplasmic space of *E. coli*,
the most common host bacterium for protein overexpression, disulfide bonds are
usually not formed, potentially making it difficult to prepare large quantities of
engineered molecules.

Thus, there is an on-going need for small, single-chain artificial

10   antibodies for a variety of therapeutic, diagnostic and catalytic applications.

SUMMARY OF THE INVENTION

The invention provides a fibronectin type III (Fn3) polypeptide
monobody comprising a plurality of Fn3 β-strand domain sequences that are
linked to a plurality of loop region sequences. One or more of the monobody

15   loop region sequences of the Fn3 polypeptide vary by deletion, insertion or
replacement of at least two amino acids from the corresponding loop region
sequences in wild-type Fn3. The β-strand domains of the monobody have at
least about 50% total amino acid sequence homology to the corresponding amino
acid sequence of wild-type Fn3's β-strand domain sequences. Preferably, one or

20   more of the loop regions of the monobody comprise amino acid residues:

    i)   from 15 to 16 inclusive in an AB loop;

    ii)   from 22 to 30 inclusive in a BC loop;

    iii) from 39 to 45 inclusive in a CD loop;

    iv)   from 51 to 55 inclusive in a DE loop;

25   v)   from 60 to 66 inclusive in an EF loop; and

    vi)   from 76 to 87 inclusive in an FG loop.

The invention also provides a nucleic acid molecule encoding a Fn3
polypeptide monobody of the invention, as well as an expression vector
comprising said nucleic acid molecule and a host cell comprising said vector.

30   The invention further provides a method of preparing a Fn3 polypeptide
monobody. The method comprises providing a DNA sequence encoding a

7

plurality of Fn3 β-strand domain sequences that are linked to a plurality of loop region sequences, wherein at least one loop region of said sequence contains a unique restriction enzyme site. The DNA sequence is cleaved at the unique restriction site. Then a preselected DNA segment is inserted into the restriction site. The preselected DNA segment encodes a peptide capable of binding to a specific binding partner (SBP) or a transition state analog compound (TSAC). The insertion of the preselected DNA segment into the DNA sequence yields a DNA molecule which encodes a polypeptide monobody having an insertion. The DNA molecule is then expressed so as to yield the polypeptide monobody.

Also provided is a method of preparing a Fn3 polypeptide monobody, which method comprises providing a replicatable DNA sequence encoding a plurality of Fn3 β-strand domain sequences that are linked to a plurality of loop region sequences, wherein the nucleotide sequence of at least one loop region is known. Polymerase chain reaction (PCR) primers are provided or prepared which are sufficiently complementary to the known loop sequence so as to be hybridizable under PCR conditions, wherein at least one of the primers contains a modified nucleic acid sequence to be inserted into the DNA sequence. PCR is performed using the replicatable DNA sequence and the primers. The reaction product of the PCR is then expressed so as to yield a polypeptide monobody.

The invention further provides a method of preparing a Fn3 polypeptide monobody. The method comprises providing a replicatable DNA sequence encoding a plurality of Fn3 β-strand domain sequences that are linked to a plurality of loop region sequences, wherein the nucleotide sequence of at least one loop region is known. Site-directed mutagenesis of at least one loop region is performed so as to create an insertion mutation. The resultant DNA comprising the insertion mutation is then expressed.

Further provided is a variegated nucleic acid library encoding Fn3 polypeptide monobodies comprising a plurality of nucleic acid species encoding a plurality of Fn3 β-strand domain sequences that are linked to a plurality of loop region sequences, wherein one or more of the monobody loop region sequences vary by deletion, insertion or replacement of at least two amino acids from

8

corresponding loop region sequences in wild-type Fn3, and wherein the β-strand

domains of the monobody have at least a 50% total amino acid sequence

homology to the corresponding amino acid sequence of β-strand domain

sequences of the wild-type Fn3. The invention also provides a peptide display

5   library derived from the variegated nucleic acid library of the invention.

Preferably, the peptide of the peptide display library is displayed on the surface

of a bacteriophage, e.g., a M13 bacteriophage or a fd bacteriophage, or virus.

The invention also provides a method of identifying the amino acid

sequence of a polypeptide molecule capable of binding to a specific binding

10 partner (SBP) so as to form a polypeptide:SSP complex, wherein the dissociation

constant of the said polypeptide:SBP complex is less than 10^{-6} moles/liter. The

method comprises the steps of:

a)   providing a peptide display library of the invention;

b)   contacting the peptide display library of (a) with an immobilized

15   or separable SBP;

c)   separating the peptide:SBP complexes from the free peptides;

d)   causing the replication of the separated peptides of (c) so as to

            result in a new peptide display library distinguished from that in

            (a) by having a lowered diversity and by being enriched in

20   displayed peptides capable of binding the SBP;

e)   optionally repeating steps (b), (c), and (d) with the new library of

            (d); and

f)   determining the nucleic acid sequence of the region encoding the

            displayed peptide of a species from (d) and hence deducing the

25   peptide sequence capable of binding to the SBP.

The present invention also provides a method of preparing a variegated

nucleic acid library encoding Fn3 polypeptide monobodies having a plurality of

nucleic acid species each comprising a plurality of loop regions, wherein the

species encode a plurality of Fn3 β-strand domain sequences that are linked to a

30 plurality of loop region sequences, wherein one or more of the loop region

sequences vary by deletion, insertion or replacement of at least two amino acids

9

from corresponding loop region sequences in wild-type Fn3, and wherein the β-strand domain sequences of the monobody have at least a 50% total amino acid sequence homology to the corresponding amino acid sequences of β-strand domain sequences of the wild-type Fn3, comprising the steps of

5   a)   preparing an Fn3 polypeptide monobody having a predetermined sequence;

    b)   contacting the polypeptide with a specific binding partner (SBP) so as to form a polypeptide:SSP complex wherein the dissociation constant of the said polypeptide:SBP complex is less than 10^{-6}

10   moles/liter;

    c)   determining the binding structure of the polypeptide:SBP complex by nuclear magnetic resonance spectroscopy or X-ray crystallography; and

    d)   preparing the variegated nucleic acid library, wherein the

15   variegation is performed at positions in the nucleic acid sequence which, from the information provided in (c), result in one or more polypeptides with improved binding to the SBP.

Also provided is a method of identifying the amino acid sequence of a polypeptide molecule capable of catalyzing a chemical reaction with a catalyzed

20 rate constant, k_{cat}, and an uncatalyzed rate constant, k_{uncat}, such that the ratio of k_{cat}/k_{uncat} is greater than 10. The method comprises the steps of:

    a)   providing a peptide display library of the invention;

    b)   contacting the peptide display library of (a) with an immobilized or separable transition state analog compound (TSAC)

25   representing the approximate molecular transition state of the chemical reaction;

    c)   separating the peptide:TSAC complexes from the free peptides;

    d)   causing the replication of the separated peptides of (c) so as to result in a new peptide display library distinguished from that in

30   (a) by having a lowered diversity and by being enriched in displayed peptides capable of binding the TSAC;

10

e)   optionally repeating steps (b), (c), and (d) with the new library of
        (d); and

f)   determining the nucleic acid sequence of the region encoding the
        displayed peptide of a species from (d) and hence deducing the
5   peptide sequence.

The invention also provides a method of preparing a variegated nucleic
acid library encoding Fn3 polypeptide monobodies having a plurality of nucleic
acid species each comprising a plurality of loop regions, wherein the species
encode a plurality of Fn3 β-strand domain sequences that are linked to a plurality
10 of loop region sequences, wherein one or more of the loop region sequences vary
by deletion, insertion or replacement of at least two amino acids from
corresponding loop region sequences in wild-type Fn3, and wherein the β-strand
domain sequences of the monobody have at least a 50% total amino acid
sequence homology to the corresponding amino acid sequences of β-strand
15 domain sequences of the wild-type Fn3, comprising the steps of

a)   preparing an Fn3 polypeptide monobody having a predetermined
        sequence, wherein the polypeptide is capable of catalyzing a
        chemical reaction with a catalyzed rate constant, k_{cat}, and an
        uncatalyzed rate constant, k_{uncat}, such that the ratio of k_{cat}/k_{uncat} is
20   greater than 10;

b)   contacting the polypeptide with an immobilized or separable
        transition state analog compound (TSAC) representing the
        approximate molecular transition state of the chemical reaction;

c)   determining the binding structure of the polypeptide:TSAC
25   complex by nuclear magnetic resonance spectroscopy or X-ray
        crystallography; and

d)   preparing the variegated nucleic acid library, wherein the
        variegation is performed at positions in the nucleic acid sequence
        which, from the information provided in (c), result in one or more
30   polypeptides with improved binding to or stabilization of the
        TSAC.

11

The invention also provides a kit for the performance of any of the methods of the invention. The invention further provides a composition, e.g., a polypeptide, prepared by the use of the kit, or identified by any of the methods of the invention.

5   The following abbreviations have been used in describing amino acids, peptides, or proteins: Ala, or A, Alanine; Arg, or R, Arginine; Asn or N, asparagine; Asp, or D, aspartic acid; Cysor C, cystein; Gln, or Q, glutamine; Glu, or E, glutamic acid; Gly, or G, glycine; His, or H, histidine; Ile, or I, isoleucine; Leu, or L, leucine; Lys, or K, lysine; Met, or M, methionine; Phe, or F,

10   phenylalanine; Pro, or P, proline; Ser, or S, serine; Thr, or T, threonine; Trp, or W, tryptophan; Tyr, or Y, tyrosine; Val, or V, valine.

The following abbreviations have been used in describing nucleic acids, DNA, or RNA: A, adenosine; T, thymidine; G, guanosine; C, cytosine.

BRIEF DESCRIPTION OF THE DRAWINGS

15   Figure 1. β-Strand and loop topology (A, B) and MOLSCRIPT representation (C, D; Kraulis, 1991) of the VH domain of anti-lysozyme immunoglobulin D1.3 (A, C; Bhat et al., 1994) and 10th type III domain of human fibronectin (B, D; Main et al., 1992). The locations of complementarity determining regions (CDRs, hypervariable regions) and the integrin-binding

20   Arg-Gly-Asp (RGD) sequence are indicated.

Figure 2. Amino acid sequence and restriction sites of the synthetic Fn3 gene. The residue numbering is according to Main et al. (1992). Restriction enzyme sites designed are shown above the amino acid sequence. β-Strands are denoted by underlines. The N-terminal "mq" sequence has been added for a

25   subsequent cloning into an expression vector. The His•tag (Novagen) fusion protein has an additional sequence, MGSSHHHHHHSSGLVPRGSH, preceding the Fn3 sequence shown above.

Figure 3. A, Far UV CD spectra of wild-type Fn3 at 25°C and 90°C. Fn3 (50 μM) was dissolved in sodium acetate (50 mM, pH 4.6). B, thermal

30   denaturation of Fn3 monitored at 215 nm. Temperature was increased at a rate of 1°C/min.

12

Figure 4. A, Cα trace of the crystal structure of the complex of lysozyme
(HEL) and the Fv fragment of the anti-hen egg-white lysozyme (anti-HEL)
antibody D1.3 (Bhat et al., 1994). Side chains of the residues 99-102 of VH
CDR3, which make contact with HEL, are also shown. B, Contact surface area
5   for each residue of the D1.3 VH-HEL and VH-VL interactions plotted vs.
residue number of D1.3 VH. Surface area and secondary structure were
determined using the program DSSP (Kabsh and Sander, 1983). C and D,
schematic drawings of the β-sheet structure of the F strand-loop-G strand
moieties of D1.3 VH (C) and Fn3 (D). The boxes denote residues in β-strands
10   and ovals those not in strands. The shaded boxes indicate residues of which side
chains are significantly buried. The broken lines indicate hydrogen bonds.

Figure 5. Designed Fn3 gene showing DNA and amino acid sequences.
The amino acid numbering is according to Main et al. (1992). The two loops
that were randomized in combinatorial libraries are enclosed in boxes.

15   Figure 6. Map of plasmid pAS45. Plasmid pAS45 is the expression
vector of His•tag-Fn3.

Figure 7. Map of plasmid pAS25. Plasmid pAS25 is the expression
vector of Fn3.

Figure 8. Map of plasmid pAS38. pAS38 is a phagmid vector for the
20   surface display of Fn3.

Figure 9. (Ubiquitin-1) Characterization of ligand-specific binding of
enriched clones using phage enzyme-linked immunosolvent assay (ELISA).
Microtiter plate wells were coated with ubiquitin (1 μg/well; "Ligand (+)) and·
then blocked with BSA. Phage solution in TBS containing approximately 10^{10}
25   colony forming units (cfu) was added to a well and washed with TBS. Bound
phages were detected with anti-phage antibody-POD conjugate (Pharmacia) with
Turbo-TMB (Pierce) as a substrate. Absorbance was measured using a
Molecular Devices SPECTRAmax 250 microplate spectrophotometer. For a
control, wells without the immobilized ligand were used. 2-1 and 2-2 denote
30   enriched clones from Library 2 eluted with free ligand and acid, respectively. 4-

1 and 4-2 denote enriched clones from Library 4 eluted with free ligand and acid, respectively.

Figure 10. (Ubiquitin-2) Competition phage ELISA of enriched clones. Phage solutions containing approximately 10^{10} cfu were first incubated with free

5   ubiquitin at 4°C for 1 hour prior to the binding to a ligand-coated well. The wells were washed and phages detected as described above.

Figure 11. Competition phage ELISA of ubiquitin-binding monobody 411. Experimental conditions are the same as described above for ubiquitin. The ELISA was performed in the presence of free ubiquitin in the binding

10 solution. The experiments were performed with four different preparations of the same clone.

Figure 12. (Fluorescein-1) Phage ELISA of four clones, pLB25.1, pLB25.4, pLB24.1 and pLB24.3. Experimental conditions are the same as ubiquitin-1 above.

15 Figure 13. (Fluorescein-2) Competition ELISA of the four clones. Experimental conditions are the same as ubiquitin-2 above.

Figure 14. ^{1}H, ^{15}N-HSQC spectrum of a fluorescence-binding monobody LB25.5. Approximately 20 µM protein was dissolved in 10 mm sodium acetate buffer (pH 5.0) containing 100 mM sodium chloride. The spectrum was

20 collected at 30°C on a Varian Unity INOVA 600 NMR spectrometer.

Figure 15. Characterization of the binding reaction of Ubi4-Fn3 to the target, ubiquitin. (a) Phage ELISA analysis of binding of Ubi4-Fn3 to ubiquitin. The binding of Ubi4-phages to ubiquitin-coated wells was measured. The control experiment was performed with wells containing no ubiquitin.

25 (b) Competition phage ELISA of Ubi4-Fn3. Ubi4-Fn3-phages were preincubated with soluble ubiquitin at an indicated concentration, followed by the phage ELISA detection in ubiquitin-coated wells.

(c) Competition phage ELISA testing the specificity of the Ubi4 clone. The Ubi4 phages were preincubated with 250 µg/ml of soluble proteins,

30 followed by phage ELISA as in (b).

(d) ELISA using free proteins.

Figure 16. Equilibrium unfolding curves for Ubi4-Fn3 (closed symbols) and wild-type Fn3 (open symbols). Squares indicate data measured in TBS (Tris HCl buffer (50 mM, pH 7.5) containing NaCl (150 mM)). Circles indicate data measured in Gly HCl buffer (20 mM, pH 3.3) containing NaCl (300 mM). The

5   curves show the best fit of the transition curve based on the two-state model. Parameters characterizing the transitions are listed in Table 7.

Figure 17. (a) ^{1}H, ^{15}N-HSQC spectrum of [^{15}N]-Ubi4-K Fn3.

(b). Difference ($\delta_{wild-type} - \delta_{Ubi4}$) of ^{1}H (b) and ^{15}N (c) chemical shifts plotted versus residue number. Values for residues 82-84 (shown as filled circles) where

10   Ubi4-K deletions are set to zero. Open circles indicate residues that are mutated in the Ubi4-K protein. The locations of β-strands are indicated with arrows.

DETAILED DESCRIPTION OF THE INVENTION

For the past decade the immune system has been exploited as a rich source of *de novo* catalysts. Catalytic antibodies have been shown to have

15   chemoselectivity, enantioselectivity, large rate accelerations, and even an ability to reroute chemical reactions. In most cases the antibodies have been elicited to transition state analog (TSA) haptens. These TSA haptens are stable, low-molecular weight compounds designed to mimic the structures of the energetically unstable transition state species that briefly (approximate half-life

20   10^{-13} s) appear along reaction pathways between reactants and products. Anti-TSA antibodies, like natural enzymes, are thought to selectively bind and stabilize transition state, thereby easing the passage of reactants to products. Thus, upon binding, the antibody lowers the energy of the actual transition state and increases the rate of the reaction. These catalysts can be programmed to

25   bind to geometrical and electrostatic features of the transition state so that the reaction route can be controlled by neutralizing unfavorable charges, overcoming entropic barriers, and dictating stereoelectronic features of the reaction. By this means even reactions that are otherwise highly disfavored have been catalyzed (Janda et al. 1997). Further, in many instances catalysts have been made for

30   reactions for which there are no known natural or man-made enzymes.

15

The success of any combinatorial chemical system in obtaining a particular function depends on the size of the library and the ability to access its members. Most often the antibodies that are made in an animal against a hapten that mimics the transition state of a reaction are first screened for binding to the

5   hapten and then screened again for catalytic activity. An improved method allows for the direct selection for catalysis from antibody libraries in phage, thereby linking chemistry and replication.

A library of antibody fragments can be created on the surface of filamentous phage viruses by adding randomized antibody genes to the gene that

10   encodes the phage's coat protein. Each phage then expresses and displays multiple copies of a single antibody fragment on its surface. Because each phage possesses both the surface-displayed antibody fragment and the DNA that encodes that fragment, and antibody fragment that binds to a target can be identified by amplifying the associated DNA.

15   Immunochemists use as antigens materials that have as little chemical reactivity as possible. It is almost always the case that one wishes the ultimate antibody to interact with native structures. In reactive immunization the concept is just the opposite. One immunizes with compounds that are highly reactive so that upon binding to the antibody molecule during the induction process, a

20   chemical reaction ensues. Later this same chemical reaction becomes part of the mechanism of the catalytic event. In a certain sense one is immunizing with a chemical reaction rather than a substance *per se*. Reactive immunogens can be considered as analogous to the mechanism-based inhibitors that enzymologists use except that they are used in the inverse way in that, instead of inhibiting a

25   mechanism, they induce a mechanism.

Man-made catalytic antibodies have considerable commercial potential in many different applications. Catalytic antibody-based products have been used successfully in prototype experiments in therapeutic applications, such as prodrug activation and cocaine inactivation, and in nontherapeutic applications,

30   such as biosensors and organic synthesis.

16

Catalytic antibodies are theoretically more attractive than noncatalytic antibodies as therapeutic agents because, being catalytic, they may be used in lower doses, and also because their effects are unusually irreversible (for example, peptide bond cleavage rather than binding). In therapy, purified

5 catalytic antibodies could be directly administered to a patient, or alternatively the patient's own catalytic antibody response could be elicited by immunization with an appropriate hapten. Catalytic antibodies also could be used as clinical diagnostic tools or as regioselective or stereoselective catalysts in the synthesis of fine chemicals.

10 I.   **Mutation of Fn3 loops and grafting of Ab loops onto Fn3**

An ideal scaffold for CDR grafting is highly soluble and stable. It is small enough for structural analysis, yet large enough to accommodate multiple CDRs so as to achieve tight binding and/or high specificity.

A novel strategy to generate an artificial Ab system on the framework of

15 an existing non-Ab protein was developed. An advantage of this approach over the minimization of an Ab scaffold is that one can avoid inheriting the undesired properties of Abs. Fibronectin type III domain (Fn3) was used as the scaffold. Fibronectin is a large protein which plays essential roles in the formation of extracellular matrix and cell-cell interactions; it consists of many repeats of three

20 types (I, II and III) of small domains (Baron et al., 1991). Fn3 itself is the paradigm of a large subfamily (Fn3 family or s-type Ig family) of the immunoglobulin superfamily (IgSF). The Fn3 family includes cell adhesion molecules, cell surface hormone and cytokine receptors, chaperonins, and carbohydrate-binding domains (for reviews, see Bork & Doolittle, 1992; Jones,

25 1993; Bork et al., 1994; Campbell & Spitzfaden, 1994; Harpez & Chothia, 1994).

Recently, crystallographic studies revealed that the structure of the DNA binding domains of the transcription factor NF-kB is also closely related to the Fn3 fold (Ghosh et al., 1995; Müller et al., 1995). These proteins are all

30 involved in specific molecular recognition, and in most cases ligand-binding sites are formed by surface loops, suggesting that the Fn3 scaffold is an excellent

framework for building specific binding proteins. The 3D structure of Fn3 has
been determined by NMR (Main et al., 1992) and by X-ray crystallography
(Leahy et al., 1992; Dickinson et al., 1994). The structure is best described as a
β-sandwich similar to that of Ab VH domain except that Fn3 has seven β-strands

5   instead of nine (Fig. 1). There are three loops on each end of Fn3; the positions
of the BC, DE and FG loops approximately correspond to those of CDR1, 2 and
3 of the VH domain, respectively (Fig. 1 C, D).

    Fn3 is small (~ 95 residues), monomeric, soluble and stable. It is one of
few members of IgSF that do not have disulfide bonds; VH has an interstrand

10 disulfide bond (Fig. 1 A) and has marginal stability under reducing conditions.
Fn3 has been expressed in *E. coli* (Aukhil et al., 1993). In addition, 17 Fn3
domains are present just in human fibronectin, providing important information
on conserved residues which are often important for the stability and folding (for
sequence alignment, see Main et al., 1992 and Dickinson et al., 1994). From

15 sequence analysis, large variations are seen in the BC and FG loops, suggesting
that the loops are not crucial to stability. NMR studies have revealed that the FG
loop is highly flexible; the flexibility has been implicated for the specific binding
of the 10th Fn3 to $\alpha_5\beta_1$ integrin through the Arg-Gly-Asp (RGD) motif. In the
crystal structure of human growth hormone-receptor complex (de Vos et al.,

20 1992), the second Fn3 domain of the receptor interacts with hormone via the FG
and BC loops, suggesting it is feasible to build a binding site using the two
loops.

    The tenth type III module of fibronectin has a fold similar to that of
immunoglobulin domains, with seven β strands forming two antiparallel β

25 sheets, which pack against each other (Main et al., 1992). The structure of the
type II module consists of seven β strands, which form a sandwich of two
antiparallel β sheets, one containing three strands (ABE) and the other four
strands (C'CFG) (Williams et al., 1988). The triple-stranded β sheet consists of
residues Glu-9-Thr-14 (A), Ser-17-Asp-23 (B), and Thr-56-Ser-60 (E). The

30 majority of the conserved residues contribute to the hydrophobic core, with the
invariant hydrophobic residues Trp-22 and Try-68 lying toward the N-terminal

18

and C-terminal ends of the core, respectively. The β strands are much less
flexible and appear to provide a rigid framework upon which functional, flexible
loops are built. The topology is similar to that of immunoglobulin C domains.

Gene construction and mutagenesis

5   A synthetic gene for tenth Fn3 of human fibronectin (Fig. 2) was
designed which includes convenient restriction sites for ease of mutagenesis and
uses specific codons for high-level protein expression (Gribskov et al., 1984).

The gene was assembled as follows: (1) the gene sequence was divided
into five parts with boundaries at designed restriction sites (Fig.2); (2) for each
10   part, a pair of oligonucleotides that code opposite strands and have
complementary overlaps of ~ 15 bases was synthesized; (3) the two
oligonucleotides were annealed and single strand regions were filled in using the
Klenow fragment of DNA polymerase; (4) the double-stranded oligonucleotide
was cloned into the pET3a vector (Novagen) using restriction enzyme sites at the
15   termini of the fragment and its sequence was confirmed by an Applied
Biosystems DNA sequencer using the dideoxy termination protocol provided by
the manufacturer; (5) steps 2-4 were repeated to obtain the whole gene (plasmid
pAS25) (Fig. 7).

Although the present method takes more time to assemble a gene than the
20   one-step polymerase chain reaction (PCR) method (Sandhu et al., 1992), no
mutations occurred in the gene. Mutations would likely have been introduced by
the low fidelity replication by Taq polymerase and would have required time-
consuming gene editing. The gene was also cloned into the pET15b (Novagen)
vector (pEW1). Both vectors expressed the Fn3 gene under the control of
25   bacteriophage T7 promoter (Studler et al. 1990); pAS25 expressed the 96-residue
Fn3 protein only, while pEW1 expressed Fn3 as a fusion protein with poly-
histidine peptide (His•tag). Recombinant DNA manipulations were performed
according to Molecular Cloning (Sambrook et al., 1989), unless otherwise stated.

Mutations were introduced to the Fn3 gene using either cassette
30   mutagenesis or oligonucleotide site-directed mutagenesis techniques (Deng &
Nickoloff, 1992). Cassette mutagenesis was performed using the same protocol

19

for gene construction described above; double-stranded DNA fragment coding a new sequence was cloned into an expression vector (pAS25 and/or pEW1). Many mutations can be made by combining a newly synthesized strand (coding mutations) and an oligonucleotide used for the gene synthesis. The resulting

5   genes were sequenced to confirm that the designed mutations and no other mutations were introduced by mutagenesis reactions.

Design and synthesis of Fn3 mutants with antibody CDRs

Two candidate loops (FG and BC) were identified for grafting. Antibodies with known crystal structures were examined in order to identify

10   candidates for the sources of loops to be grafted onto Fn3. Anti-hen egg lysozyme (HEL) antibody D1.3 (Bhat et al., 1994) was chosen as the source of a CDR loop. The reasons for this choice were: (1) high resolution crystal structures of the free and complexed states are available (Fig. 4 A; Bhat et al., 1994), (2) thermodynamics data for the binding reaction are available (Tello et

15   al., 1993), (3) D1.3 has been used as a paradigm for Ab structural analysis and Ab engineering (Verhoeyen et al., 1988; McCafferty et al., 1990) (4) site-directed mutagenesis experiments have shown that CDR3 of the heavy chain (VH-CDR3) makes a larger contribution to the affinity than the other CDRs (Hawkins et al., 1993), and (5) a binding assay can be easily performed. The

20   objective for this trial was to graft VH-CDR3 of D1.3 onto the Fn3 scaffold without significant loss of stability.

An analysis of the D1.3 structure (Fig. 4) revealed that only residues 99-102 ("RDYR") make direct contact with hen egg-white lysozyme (HEL) (Fig. 4 B), although VH-CDR3 is defined as longer (Bhat et al., 1994). It should be

25   noted that the C-terminal half of VH-CDR3 (residues 101-104) made significant contact with the VL domain (Fig. 4 B). It has also become clear that D1.3 VH-CDR3 (Fig. 4 C) has a shorter turn between the strands F and G than the FG loop of Fn3 (Fig. 4 D). Therefore, mutant sequences were designed by using the RDYR (99-102) of D1.3 as the core and made different boundaries and loop

30   lengths (Table 1). Shorter loops may mimic the D1.3 CDR3 conformation

better, thereby yielding higher affinity, but they may also significantly reduce stability by removing wild-type interactions of Fn3.

Table 1. Amino acid sequences of D1.3 VH CDR3, VH8 CDR3 and Fn3 FG loop and list of planned mutants.

	96	100	105
	•	•	•
D1.3	A R E R D Y R L D Y W G Q G		
VH8	A R G A V V S Y Y A M D Y W G Q G		
	75	80	85
	•	•	•
Fn3	Y A V T G R G D S P A S S K P I		

Mutant	Sequence
D1.3-1	Y A E R D Y R L D Y – – – – P I
D1.3-2	Y A V R D Y R L D Y – – – – P I
D1.3-3	Y A V R D Y R L D Y A S S K P I
D1.3-4	Y A V R D Y R L D Y – – – K P I
D1.3-5	Y A V R D Y R – – – – – S K P I
D 1.3-6	Y A V T R D Y R L – – S S K P I
D1.3-7	Y A V T E R D Y R L – S S K P I
VH8-1	Y A V A V V S Y Y A M D Y – P I
VH8-2	Y A V T A V V S Y Y A S S K P I

Underlines indicate residues in β-strands. Bold characters indicate replaced residues.

In addition, an anti-HEL single VH domain termed VH8 (Ward et al., 1989) was chosen as a template. VH8 was selected by library screening and, in spite of the lack of the VL domain, VH8 has an affinity for HEL of 27 nM, probably due to its longer VH-CDR3 (Table 1). Therefore, its VH-CDR3 was grafted onto Fn3. Longer loops may be advantageous on the Fn3 framework because they may provide higher affinity and also are close to the loop length of wild-type Fn3. The 3D structure of VH8 was not known and thus the VH8 CDR3 sequence was aligned with that of D1.3 VH-CDR3; two loops were designed (Table 1).

Mutant construction and production

Site-directed mutagenesis experiments were performed to obtain designed sequences. Two mutant Fn3s, D1.3-1 and D1.3-4 (Table 1) were obtained and both were expressed as soluble His•tag fusion proteins. D1.3-4 was

5   purified and the His•tag portion was removed by thrombin cleavage. D1.3-4 is soluble up to at least 1 mM at pH 7.2. No aggregation of the protein has been observed during sample preparation and NMR data acquisition.

Protein expression and purification

E. coli BL21 (DE3) (Novagen) were transformed with an expression

10 vector (pAS25, pEW1 and their derivatives) containing a gene for the wild-type or a mutant. Cells were grown in M9 minimal medium and M9 medium supplemented with Bactotrypton (Difco) containing ampicillin (200 µg/ml). For isotopic labeling, ^{15}N NH$_4$Cl and/or ^{13}C glucose replaced unlabeled components. 500 ml medium in a 2 liter baffle flask were inoculated with 10 ml of overnight

15 culture and agitated at 37°C. Isopropylthio-β-galactoside (IPTG) was added at a final concentration of 1 mM to initiate protein expression when OD (600 nm) reaches one. The cells were harvested by centrifugation 3 hours after the addition of IPTG and kept frozen at -70°C until used.

Fn3 without His•tag was purified as follows. Cells were suspended in

20 5 ml/(g cell) of Tris (50 mM, pH 7.6) containing ethylenediaminetetraacetic acid (EDTA; 1 mM) and phenylmethylsulfonyl fluoride (1 mM). HEL was added to a final concentration of 0.5 mg/ml. After incubating the solution for 30 minutes at 37°C, it was sonicated three times for 30 seconds on ice. Cell debris was removed by centrifugation. Ammonium sulfate was added to the solution and

25 precipitate recovered by centrifugation. The pellet was dissolved in 5-10 ml sodium acetate (50 mM, pH 4.6) and insoluble material was removed by centrifugation. The solution was applied to a Sephacryl S100HR column (Pharmacia) equilibrated in the sodium acetate buffer. Fractions containing Fn3 then was applied to a ResourceS column (Pharmacia) equilibrated in sodium

30 acetate (50 mM, pH 4.6) and eluted with a linear gradient of sodium chloride (0-

22

0.5 M). The protocol can be adjusted to purify mutant proteins with different surface charge properties.

Fn3 with His•tag was purified as follows. The soluble fraction was prepared as described above, except that sodium phosphate buffer (50 mM, pH

5   7.6) containing sodium chloride (100 mM) replaced the Tris buffer. The solution was applied to a Hi-Trap chelating column (Pharmacia) preloaded with nickel and equilibrated in the phosphate buffer. After washing the column with the buffer, His•tag-Fn3 was eluted in the phosphate buffer containing 50 mM EDTA. Fractions containing His•tag-Fn3 were pooled and applied to a

10 Sephacryl S100-HR column, yielding highly pure protein. The His•tag portion was cleaved off by treating the fusion protein with thrombin using the protocol supplied by Novagen. Fn3 was separated from the His•tag peptide and thrombin by a ResourceS column using the protocol above.

The wild-type and two mutant proteins so far examined are expressed as

15 soluble proteins. In the case that a mutant is expressed as inclusion bodies (insoluble aggregate), it is first examined if it can be expressed as a soluble protein at lower temperature (e.g., 25-30°C). If this is not possible, the inclusion bodies are collected by low-speed centrifugation following cell lysis as described above. The pellet is washed with buffer, sonicated and centrifuged. The

20 inclusion bodies are solubilized in phosphate buffer (50 mM, pH 7.6) containing guanidinium chloride (GdnCl, 6 M) and will be loaded on a Hi-Trap chelating column. The protein is eluted with the buffer containing GdnCl and 50 mM EDTA.

Conformation of mutant Fn3, D1.3-4

25 The ^{1}H NMR spectra of His•tag D1.3-4 fusion protein closely resembled that of the wild-type, suggesting the mutant is folded in a similar conformation to that of the wild-type. The spectrum of D1.3-4 after the removal of the His•tag peptide showed a large spectral dispersion. A large dispersion of amide protons (7-9.5 ppm) and a large number of downfield (5.0-6.5 ppm) C^α protons are

30 characteristic of a β-sheet protein (Wüthrich, 1986).

23

    The 2D NOESY spectrum of D1.3-4 provided further evidence for a preserved conformation. The region in the spectrum showed interactions between upfield methyl protons (< 0.5 ppm) and methyl-methylene protons. The Val72 γ methyl resonances were well separated in the wild-type spectrum (-0.07

5   and 0.37 ppm; (Baron et al., 1992)). Resonances corresponding to the two methyl protons are present in the D1.3-4 spectrum (-0.07 and 0.44 ppm). The cross peak between these two resonances and other conserved cross peaks indicate that the two resonances in the D1.3-4 spectrum are highly likely those of Val72 and that other methyl protons are in nearly identical environment to that

10   of wild-type Fn3. Minor differences between the two spectra are presumably due to small structural perturbation due to the mutations. Val72 is on the F strand, where it forms a part of the central hydrophobic core of Fn3 (Main et al., 1992). It is only four residues away from the mutated residues of the FG loop (Table 1). The results are remarkable because, despite there being 7 mutations

15   and 3 deletions in the loop (more than 10% of total residues; Fig. 12, Table 2), D1.3-4 retains a 3D structure virtually identical to that of the wild-type (except for the mutated loop). Therefore, the results provide strong support that the FG loop is not significantly contributing to the folding and stability of the Fn3 molecule and thus that the FG loop can be mutated extensively.

20

Table 2. Sequences of oligonucleotides

Name	Sequence
FN1F	CGGGATCC<u>CATATG</u>CAGGTTTCTGATGTTCCGCGTGACCTGGAAGTTGTTGCTGCGACC
FN1R	TAA<u>CTGCAG</u>GAGCATCCCAGCTGATCAGCAGGCTAGTCGGGGTCGCAGCAACAAC
FN2F	CTC<u>CTGCAG</u>TTACCGTGCGTTATTACCGTATCACGTACGGTGAAACCGGTG
FN2R	GT<u>GAATTC</u>CTGAACCGGGGAGTTACCACCGGTTTCACCG
FN3F	AG<u>GAATTC</u>ACTGTACCTGGTTCCAAGTCTACTGCTACCATCAGCGG
FN3R	GTATA<u>GTCGAC</u>ACCCGGTTTCAGGCCGCTGATGGTAGC
FN4F	CGGGT<u>GTCGAC</u>TATACCATCACTGTATACGCT
FN4R	CGGGATCC<u>GAGCTC</u>GCTGGGCTGTCACCACGGCCAGTAACAGCGTATACAGTGAT
FN5F	CAGC<u>GAGCTC</u>CAAGCCAATCTCGATTAACTACCGT
FN5R	CG<u>GGATCC</u>TCGAGTTACTAGGTACGGTAGTTAATCGA
FN5R'	CG<u>GGATCC</u>ACGCGTGCCACCGGTACGGTAGTTAATCGA
gene3F	CG<u>GGATCC</u>ACGCGTCCATTCGTTTGTGAATATCAAGGCCAATCG

24

```
gene3R   CCGGAAGCTTTAAGACTCCTTATTACGCAGTATGTTAGC
38TAABgIII        CTGTTACTGGCCGTGAGATCTAACCAGCGAGCTCCA
BC3      GATCAGCTGGGATGCTCCTNNKNNKNNKNNKNNKTATTACCGTATCACGTA
FG2      TGTATACGCTGTTACTGGCNNKNNKNNKNNKNNKNNKNNKTCCAAGCCAATCTCGAT
FG3      CTGTATACGCTGTTACTGGCNNKNNKNNKNNKCCAGCGAGCTCCAAG
FG4      CATCACTGTATACGCTGTTACTNNKNNKNNKNNKNNKTCCAAGCCAATCTC
```

Restriction enzyme sites are underlined. N and K denote an equimolar mixture of A, T. G and C and that of G and T, respectively.

Structure and stability measurements

Structures of Abs were analyzed using quantitative methods (e.g., DSSP (Kabsch & Sander, 1983) and PDBfit (D. McRee, The Scripps Research Institute)) as well as computer graphics (e.g., Quanta (Molecular Simulations) and What if (G. Vriend, European Molecular Biology Laboratory)) to superimpose the strand-loop-strand structures of Abs and Fn3.

The stability of FnAbs was determined by measuring temperature- and chemical denaturant-induced unfolding reactions (Pace et al., 1989). The temperature-induced unfolding reaction was measured using a circular dichroism (CD) polarimeter. Ellipticity at 222 and 215 nm was recorded as the sample temperature was slowly raised. Sample concentrations between 10 and 50 μM were used. After the unfolding baseline was established, the temperature was lowered to examine the reversibility of the unfolding reaction. Free energy of unfolding was determined by fitting data to the equation for the two-state transition (Becktel & Schellman, 1987; Pace et al., 1989). Nonlinear least-squares fitting was performed using the program Igor (WaveMetrics) on a Macintosh computer.

The structure and stability of two selected mutant Fn3s were studied; the first mutant was D1.3-4 (Table 2) and the second was a mutant called AS40 which contains four mutations in the BC loop ($A^{26}V^{27}T^{28}V^{29}$) → TQRQ). AS40 was randomly chosen from the BC loop library described above. Both mutants were expressed as soluble proteins in *E. coli* and were concentrated at least to 1 mM, permitting NMR studies.

25

The mid-point of the thermal denaturation for both mutants was approximately 69°C, as compared to approximately 79°C for the wild-type protein. The results indicated that the extensive mutations at the two surface loops did not drastically decrease the stability of Fn3, and thus demonstrated the

5   feasibility of introducing a large number of mutations in both loops.

Stability was also determined by guanidinium chloride (GdnCl)- and urea-induced unfolding reactions. Preliminary unfolding curves were recorded using a fluorometer equipped with a motor-driven syringe; GdnCl or urea were added continuously to the protein solution in the cuvette. Based on the

10   preliminary unfolding curves, separate samples containing varying concentration of a denaturant were prepared and fluorescence (excitation at 290 nm, emission at 300-400 nm) or CD (ellipticity at 222 and 215 nm) were measured after the samples were equilibrated at the measurement temperature for at least one hour. The curve was fitted by the least-squares method to the equation for the two-state

15   model (Santoro & Bolen, 1988; Koide et al., 1993). The change in protein concentration was compensated if required.

Once the reversibility of the thermal unfolding reaction is established, the unfolding reaction is measured by a Microcal MC-2 differential scanning calorimeter (DSC). The cell (~ 1.3 ml) will be filled with FnAb solution (0.1 -

20   1 mM) and ΔCp (= $\Delta H/\Delta T$) will be recorded as the temperature is slowly raised. T_m (the midpoint of unfolding), ΔH of unfolding and ΔG of unfolding is determined by fitting the transition curve (Privalov & Potekhin, 1986) with the Origin software provided by Microcal.

Thermal unfolding

25   A temperature-induced unfolding experiment on Fn3 was performed using circular dichroism (CD) spectroscopy to monitor changes in secondary structure. The CD spectrum of the native Fn3 shows a weak signal near 222 nm (Fig. 3A), consistent with the predominantly β-structure of Fn3 (Perczel et al., 1992). A cooperative unfolding transition is observed at 80-90°C, clearly

30   indicating high stability of Fn3 (Fig. 3B). The free energy of unfolding could not be determined due to the lack of a post-transition baseline. The result is

26

consistent with the high stability of the first Fn3 domain of human fibronectin (Litvinovich et al., 1992), thus indicating that Fn3 domains are in general highly stable.

Binding assays

5   Binding reaction of FnAbs were characterized quantitatively using an isothermal titration calorimeter (ITC) and fluorescence spectroscopy.

The enthalpy change (ΔH) of binding were measured using a Microcal Omega ITC (Wiseman et al., 1989). The sample cell (~ 1.3 ml) was filled with FnAbs solution ($\leq$ 100 μM, changed according to K_d), and the reference cell

10   filled with distilled water; the system was equilibrated at a given temperature until a stable baseline is obtained; 5-20 μl of ligand solution ($\leq$ 2 mM) was injected by a motor-driven syringe within a short duration (20 sec) followed by an equilibration delay (4 minutes); the injection was repeated and heat generation/absorption for each injection was measured. From the change in the

15   observed heat change as a function of ligand concentration, ΔH and K_d was determined (Wiseman et al., 1989). ΔG and ΔS of the binding reaction was deduced from the two directly measured parameters. Deviation from the theoretical curve was examined to assess nonspecific (multiple-site) binding. Experiments were also be performed by placing a ligand in the cell and titrating

20   with an FnAb. It should be emphasized that only ITC gives direct measurement of ΔH, thereby making it possible to evaluate enthalpic and entropic contributions to the binding energy. ITC was successfully used to monitor the binding reaction of the D1.3 Ab (Tello et al., 1993; Bhat et al., 1994).

Intrinsic fluorescence is monitored to measure binding reactions with K_d

25   in the sub-μM range where the determination of K_d by ITC is difficult. Trp fluorescence (excitation at ~ 290 nm, emission at 300-350 nm) and Tyr fluorescence (excitation at ~ 260 nm, emission at ~ 303 nm) is monitored as the Fn3-mutant solution ($\leq$ 10 μM) is titrated with ligand solution ($\leq$ 100 μM). K_d of the reaction is determined by the nonlinear least-squares fitting of the

30   bimolecular binding equation. Presence of secondary binding sites is examined using Scatchard analysis. In all binding assays, control experiments are

27

performed busing wild-type Fn3 (or unrelated FnAbs) in place of FnAbs of
interest.

II.   Production of Fn3 mutants with high affinity and specificity FnAbs

Library screening was carried out in order to select FnAbs which bind to
5   specific ligands. This is complementary to the modeling approach described
above. The advantage of combinatorial screening is that one can easily produce
and screen a large number of variants ($\geq 10^8$), which is not feasible with specific
mutagenesis ("rational design") approaches. The phage display technique
(Smith, 1985; O'Neil & Hoess, 1995) was used to effect the screening processes.
10   Fn3 was fused to a phage coat protein (pIII) and displayed on the surface of
filamentous phages. These phages harbor a single-stranded DNA genome that
contains the gene coding the Fn3 fusion protein. The amino acid sequence of
defined regions of Fn3 were randomized using a degenerate nucleotide sequence,
thereby constructing a library. Phages displaying Fn3 mutants with desired
15   binding capabilities were selected *in vitro*, recovered and amplified. The amino
acid sequence of a selected clone can be identified readily by sequencing the Fn3
gene of the selected phage. The protocols of Smith (Smith & Scott, 1993) were
followed with minor modifications.

The objective was to produce FnAbs which have high affinity to small
20   protein ligands. HEL and the B1 domain of staphylococcal protein G (hereafter
referred to as protein G) were used as ligands. Protein G is small (56 amino
acids) and highly stable (Minor & Kim, 1994; Smith et al., 1994). Its structure
was determined by NMR spectroscopy (Gronenborn et al., 1991) to be a helix
packed against a four-strand β-sheet. The resulting FnAb-protein G complexes
25   (~ 150 residues) is one of the smallest protein-protein complexes produced to
date, well within the range of direct NMR methods. The small size, the high
stability and solubility of both components and the ability to label each with
stable isotopes (^{13}C and ^{15}N; see below for protein G) make the complexes an
ideal model system for NMR studies on protein-protein interactions.
30   The successful loop replacement of Fn3 (the mutant D1.3-4) demonstrate
that at least ten residues can be mutated without the loss of the global fold.

28

Based on this, a library was first constructed in which only residues in the FG
loop are randomized. After results of loop replacement experiments on the BC
loop were obtained, mutation sites were extended that include the BC loop and
other sites.

5   **Construction of Fn3 phage display system**

An M13 phage-based expression vector pASM1 has been constructed as
follows: an oligonucleotide coding the signal peptide of OmpT was cloned at
the 5' end of the Fn3 gene; a gene fragment coding the C-terminal domain of
M13 pIII was prepared from the wild-type gene III gene of M13 mp18 using

10   PCR (Corey et al., 1993) and the fragment was inserted at the 3' end of the
OmpT-Fn3 gene; a spacer sequence has been inserted between Fn3 and pIII. The
resultant fragment (OmpT-Fn3-pIII) was cloned in the multiple cloning site of
M13 mp18, where the fusion gene is under the control of the lac promoter. This
system will produce the Fn3-pIII fusion protein as well as the wild-type pIII

15   protein. The co-expression of wild-type pIII is expected to reduce the number of
fusion pIII protein, thereby increasing the phage infectivity (Corey et al., 1993)
(five copies of pIII are present on a phage particle). In addition, a smaller
number of fusion pIII protein may be advantageous in selecting tight binding
proteins, because the chelating effect due to multiple binding sites should be

20   smaller than that with all five copies of fusion pIII (Bass et al., 1990). This
system has successfully displayed the serine protease trypsin (Corey et al.,
1993). Phages were produced and purified using *E. coli* K91kan (Smith & Scott,
1993) according to a standard method (Sambrook et al., 1989) except that phage
particles were purified by a second polyethylene glycol precipitation and acid

25   precipitation.

Successful display of Fn3 on fusion phages has been confirmed by
ELISA using an Ab against fibronectin (Sigma), clearly indicating that it is
feasible to construct libraries using this system.

An alternative system using the fUSE5 (Parmley & Smith, 1988) may

30   also be used. The Fn3 gene is inserted to fUSE5 using the SfiI restriction sites
introduced at the 5'- and 3'- ends of the Fn3 gene PCR. This system displays

only the fusion pIII protein (up to five copies) on the surface of a phage. Phages are produced and purified as described (Smith & Scott, 1993). This system has been used to display many proteins and is robust. The advantage of fUSE5 is its low toxicity. This is due to the low copy number of the replication form (RF) in the host, which in turn makes it difficult to prepare a sufficient amount of RF for library construction (Smith & Scott, 1993).

Construction of libraries

The first library was constructed of the Fn3 domain displayed on the surface of MB phage in which seven residues (77-83) in the FG loop (Fig. 4D) were randomized. Randomization will be achieved by the use of an oligonucleotide containing degenerated nucleotide sequence. A double-stranded nucleotide was prepared by the same protocol as for gene synthesis (see above) except that one strand had an $(NNK)_6(NNG)$ sequence at the mutation sites, where N corresponds to an equimolar mixture of A, T, G and C and K corresponds to an equimolar mixture of G and T. The (NNG) codon at residue 83 was required to conserve the SacI restriction site (Fig. 2). The (NNK) codon codes all of the 20 amino acids, while the NNG codon codes 14. Therefore, this library contained ~ 10^9 independent sequences. The library was constructed by ligating the double-stranded nucleotide into the wild-type phage vector, pASM1, and the transfecting *E. coli* XL1 blue (Stratagene) using electroporation. XL1 blue has the lacI^q phenotype and thus suppresses the expression of the Fn3-pIII fusion protein in the absence of lac inducers. The initial library was propagated in this way, to avoid selection against toxic Fn3-pIII clones. Phages displaying the randomized Fn3-pIII fusion protein were prepared by propagating phages with K91kan as the host. K91kan does not suppress the production of the fusion protein, because it does not have lacI^q. Another library was also generated in which the BC loop (residues 26-20) was randomized.

Selection of displayed FnAbs

Screening of Fn3 phage libraries was performed using the biopanning protocol (Smith & Scott, 1993); a ligand is biotinylated and the strong biotin-streptavidin interaction was used to immobilize the ligand on a streptavidin-

coated dish. Experiments were performed at room temperature (~ 22°C). For
the initial recovery of phages from a library, 10 μg of a biotinylated ligand were
immobilized on a streptavidin-coated polystyrene dish (35 mm, Falcon 1008)
and then a phage solution (containing ~ 10^{11} pfu (plaque-forming unit)) was

5   added. After washing the dish with an appropriate buffer (typically TBST, Tris-
HCl (50 mM, pH 7.5), NaCl (150 mM) and Tween 20 (0.5%)), bound phages
were eluted by one or combinations of the following conditions: low pH, an
addition of a free ligand, urea (up to 6 M) and, in the case of anti-protein G
FnAbs, cleaving the protein G-biotin linker by thrombin. Recovered phages

10   were amplified using the standard protocol using K91kan as the host (Sambrook
et al., 1989). The selection process were repeated 3-5 times to concentrate
positive clones. From the second round on, the amount of the ligand were
gradually decreased (to ~ 1 μg) and the biotinylated ligand were mixed with a
phage solution before transferring a dish (G. P. Smith, personal communication).

15   After the final round, 10-20 clones were picked, and their DNA sequence will be
determined. The ligand affinity of the clones were measured first by the phage-
ELISA method (see below).

To suppress potential binding of the Fn3 framework (background
binding) to a ligand, wild-type Fn3 may be added as a competitor in the buffers.

20   In addition, unrelated proteins (e.g., bovine serum albumin, cytochrome c and
RNase A) may be used as competitors to select highly specific FnAbs.

Binding assay

The binding affinity of FnAbs on phage surface is characterized semi-
quantitatively using the phage ELISA technique (Li et al., 1995). Wells of

25   microtiter plates (Nunc) are coated with a ligand protein (or with streptavidin
followed by the binding of a biotinylated ligand) and blocked with the Blotto
solution (Pierce). Purified phages (~ 10^{10} pfu) originating from single plaques
(M13)/colonies (fUSE5) are added to each well and incubated overnight at 4°C.
After washing wells with an appropriate buffer (see above), bound phages are

30   detected by the standard ELISA protocol using anti-M13 Ab (rabbit, Sigma) and
anti-rabbit Ig-peroxidase conjugate (Pierce) or using anti-M13 Ab-peroxidase

conjugate (Pharmacia). Colormetric assays are performed using TMB (3,3',5,5'-tetramethylbenzidine, Pierce). The high affinity of protein G to immunoglobulins present a special problem; Abs cannot be used in detection. Therefore, to detect anti-protein G FnAbs, fusion phages are immobilized in wells and the binding is then measured using biotinylated protein G followed by the detection using streptavidin-peroxidase conjugate.

Production of soluble FnAbs

After preliminary characterization of mutant Fn3s using phage ELISA, mutant genes are subcloned into the expression vector pEW1. Mutant proteins are produced as His•tag fusion proteins and purified, and their conformation, stability and ligand affinity are characterized.

Thus, Fn3 is the fourth example of a monomeric immunoglobulin-like scaffold that can be used for engineering binding proteins. Successful selection of novel binding proteins have also been based on minibody, tendamistat and "camelized" immunoglobulin VH domain scaffolds (Martin et al., 1994; Davies & Riechmann, 1995; McConnell & Hoess, 1995). The Fn3 scaffold has advantages over these systems. Bianchi et al. reported that the stability of a minibody was 2.5 kcal/mol, significantly lower than that of Ubi4-K. No detailed structural characterization of minibodies has been reported to date. Tendamistat and the VH domain contain disulfide bonds, and thus preparation of correctly folded proteins may be difficult. Davies and Riechmann reported that the yields of their camelized VH domains were less than 1 mg per liter culture (Davies & Riechmann, 1996).

Thus, the Fn3 framework can be used as a scaffold for molecular recognition. Its small size, stability and well-characterized structure make Fn3 an attractive system. In light of the ubiquitous presence of Fn3 in a wide variety of natural proteins involved in ligand binding, one can engineer Fn3-based binding proteins to different classes of targets.

The following examples are intended to illustrate but not limit the invention.

32

EXAMPLE I

Construction of the Fn3 gene

A synthetic gene for tenth Fn3 of fibronectin (Fig.1) was designed on the basis of amino acid residue 1416-1509 of human fibronectin (Kornblihtt, *et al.*,

5   1985) and its three dimensional structure (Main, *et al.*, 1992). The gene was engineered to include convenient restriction sites for mutagenesis and the so-called "preferred codons" for high level protein expression (Gribskov, *et al.*, 1984) were used. In addition, a glutamine residue was inserted after the N-terminal methionine in order to avoid partial processing of the N-terminal

10   methionine which often degrades NMR spectra (Smith, *et al.*, 1994). Chemical reagents were of the analytical grade or better and purchased from Sigma Chemical Company and J.T. Baker, unless otherwise noted. Recombinant DNA procedures were performed as described in "Molecular Cloning" (Sambrook, *et al.*, 1989), unless otherwise stated. Custom oligonucleotides were purchased

15   from Operon Technologies. Restriction and modification enzymes were from New England Biolabs.

The gene was assembled in the following manner. First, the gene sequence (Fig. 5) was divided into five parts with boundaries at designed restriction sites: fragment 1, NdeI-PstI (oligonucleotides FN1F and FN1R (Table

20   2); fragment 2, PstI-EcoRI (FN2F and FN2R); fragment 3, EcoRI-SalI (FN3F and FN3R); fragment 4, SalI-SacI (FN4F and FN4R); fragment 5, SacI-BamHI (FN5F and FN5R). Second, for each part, a pair of oligonucleotides which code opposite strands and have complementary overlaps of approximately 15 bases was synthesized. These oligonucleotides were designated FN1F-FN5R and are

25   shown in Table 2. Third, each pair (e.g., FN1F and FN1R) was annealed and single-strand regions were filled in using the Klenow fragment of DNA polymerase. Fourth, the double stranded oligonucleotide was digested with the relevant restriction enzymes at the termini of the fragment and cloned into the pBlueScript SK plasmid (Stratagene) which had been digested with the same

30   enzymes as those used for the fragments. The DNA sequence of the inserted fragment was confirmed by DNA sequencing using an Applied Biosystems DNA

sequencer and the dideoxy termination protocol provided by the manufacturer.

Last, steps 2-4 were repeated to obtain the entire gene.

The gene was also cloned into the pET3a and pET15b (Novagen) vectors

(pAS45 and pAS25, respectively). The maps of the plasmids are shown in Figs.

5   6 and 7. *E. coli* BL21 (DE3) (Novagen) containing these vectors expressed the

Fn3 gene under the control of bacteriophage T7 promotor (Studier, *et al.*, 1990);

pAS24 expresses the 96-residue Fn3 protein only, while pAS45 expresses Fn3 as

a fusion protein with poly-histidine peptide (His•tag). High level expression of

the Fn3 protein and its derivatives in *E. coli* was detected as an intense band on

10 SDS-PAGE stained with CBB.

The binding reaction of the monobodies is characterized quantitatively by

means of fluorescence spectroscopy using purified soluble monobodies.

Intrinsic fluorescence is monitored to measure binding reactions. Trp

fluorescence (excitation at ~290 nm, emission at 300 350 nm) and Tyr

15 fluorescence (excitation at ~260 nm, emission at ~303 nm) is monitored as the

Fn3-mutant solution ($\leq$ 100 μM) is titrated with a ligand solution. When a

ligand is fluorescent (e.g. fluorescein), fluorescence from the ligand may be

used. K_d of the reaction will be determined by the nonlinear least-squares fitting

of the bimolecular binding equation.

20 If intrinsic fluorescence cannot be used to monitor the binding reaction,

monobodies are labeled with fluorescein-NHS (Pierce) and fluorescence

polarization is used to monitor the binding reaction (Burke *et al.*, 1996).

EXAMPLE II

Modifications to include restriction sites in the Fn3 gene

25 The restriction sites were incorporated in the synthetic Fn3 gene without

changing the amino acid sequence Fn3. The positions of the restriction sites

were chosen so that the gene construction could be completed without

synthesizing long (>60 bases) oligonucleotides and so that two loop regions

could be mutated (including by randomization) by the cassette mutagenesis

30 method (i.e., swapping a fragment with another synthetic fragment containing

mutations). In addition, the restriction sites were chosen so that most sites were

34

unique in the vector for phage display. Unique restriction sites allow one to
recombine monobody clones which have been already selected in order to supply
a larger sequence space.

EXAMPLE III

5   ### Construction of M13 phage display libraries

A vector for phage display, pAS38 (for its map, see Fig. 8) was
constructed as follows. The XbaI-BamHI fragment of pET12a encoding the
signal peptide of OmpT was cloned at the 5' end of the Fn3 gene. The C-
terminal region (from the FN5F and FN5R oligonucleotides, see Table 2) of the
10   Fn3 gene was replaced with a new fragment consisting of the FN5F and FN5R'
oligonucleotides (Table 2) which introduced a MluI site and a linker sequence
for making a fusion protein with the pIII protein of bacteriophage M13. A gene
fragment coding the C-terminal domain of M13 pIII was prepared from the wild-
type gene III of M13mp18 using PCR (Corey, *et al.*, 1993) and the fragment was
15   inserted at the 3' end of the OmpT-Fn3 fusion gene using the MluI and HindIII
sites.

Phages were produced and purified using a helper phage, M13K07,
according to a standard method (Sambrook, *et al.*, 1989) except that phage
particles were purified by a second polyethylene glycol precipitation. Successful
20   display of Fn3 on fusion phages was confirmed by ELISA (Harlow & Lane,
1988) using an antibody against fibronectin (Sigma) *and a custom anti-FN3
antibody (Cocalico Biologicals, PA, USA).*

EXAMPLE IV

Libraries containing loop variegations in the AB loop

25   A nucleic acid phage display library having variegation in the AB loop is
prepared by the following methods. Randomization is achieved by the use of
oligonucleotides containing degenerated nucleotide sequence. Residues to be
variegated are identified by examining the X-ray and NMR structures of Fn3
(Protein Data Bank accession numbers, 1FNA and 1TTF, respectively).

30   Oligonucleotides containing NNK (N and K here denote an equimolar mixture of
A, T, G, and C and an equimolar mixture of G and T, respectively) for the

35

variegated residues are synthesized (see oligonucleotides BC3, FG2, FG3, and FG4 in Table 2 for example). The NNK mixture codes for all twenty amino acids and one termination codon (TAG). TAG, however, is suppressed in the *E. coli* XL-1 blue. Single-stranded DNAs of pAS38 (and its derivatives) are

5   prepared using a standard protocol (Sambrook, *et al.*, 1989).

Site-directed mutagenesis is performed following published methods (see for example, Kunkel, 1985) using a Muta-Gene kit (BioRad). The libraries are constructed by electroporation of *E. coli* XL-1 Blue electroporation competent cells (200 µl; Stratagene) with 1 µg of the plasmid DNA using a BTX electrocell

10 manipulator ECM 395 1mm gap cuvette. A portion of the transformed cells is plated on an LB-agar plate containing ampicillin (100 µg/ml) to determine the transformation efficiency. Typically, 3×10^8 transformants are obtained with 1 µg of DNA, and thus a library contains 10^8 to 10^9 independent clones. Phagemid particles were prepared as described above.

15   **EXAMPLE V**

Loop variegations in the BC, CD, DE, EF or FG loop

A nucleic acid phage display library having five variegated residues (residues number 26-30) in the BC loop, and one having seven variegated residues (residue numbers 78-84) in the FG loop, was prepared using the

20 methods described in Example IV above. Other nucleic acid phage display libraries having variegation in the CD, DE or EF loop can be prepared by similar methods.

EXAMPLE VI

Loop variegations in the FG and BC loop

25   A nucleic acid phage display library having seven variegated residues (residues number 78-84) in the FG loop and five variegated residues (residue number 26-30) in the BC loop was prepared. Variegations in the BC loop were prepared by site-directed mutagenesis (Kunkel, *et al.*) using the BC3 oligonucleotide described in Table 1. Variegations in the FG loop were

30 introduced using site-directed mutagenesis using the BC loop library as the starting material, thereby resulting in libraries containing variegations in both

36

BC and FG loops. The oligonucleotide FG2 has variegating residues 78-84 and

oligonucleotide FG4 has variegating residues 77-81 and a deletion of residues

82-84.

A nucleic acid phage display library having five variegated residues

5   (residues 78-84) in the FG loop and a three residue deletion (residues 82-84) in

the FG loop, and five variegated residues (residues 26-30) in the BC loop, was

prepared. The shorter FG loop was made in an attempt to reduce the flexibility of

the FG loop; the loop was shown to be highly flexible in Fn3 by the NMR

studies of Main, *et al.* (1992). A highly flexible loop may be disadvantageous to

10   forming a binding site with a high affinity (a large entropy loss is expected upon

the ligand binding, because the flexible loop should become more rigid). In

addition, other Fn3 domains (besides human) have shorter FG loops (for

sequence alignment, see Figure 12 in Dickinson, *et al.* (1994)).

Randomization was achieved by the use of oligonucleotides containing

15   degenerate nucleotide sequence (oligonucleotide BC3 for variegating the BC

loop and oligonucleotides FG2 and FG4 for variegating the FG loops).

Site-directed mutagenesis was performed following published methods

(see for example, Kunkel, 1985). The libraries were constructed by

electrotransforming *E. coli* XL-1 Blue (Stratagene). Typically a library contains

20   10^8 to 10^9 independent clones. Library 2 contains five variegated residues in the

BC loop and seven variegated residues in the FG loop. Library 4 contains five

variegated residues in each of the BC and FG loops, and the length of the FG

loop was shortened by three residues.

EXAMPLE VII

25   **fd phage display libraries constructed with loop variegations**

Phage display libraries are constructed using the fd phage as the genetic

vector. The Fn3 gene is inserted in fUSE5 (Parmley & Smith, 1988) using SfiI

restriction sites which are introduced at the 5' and 3' ends of the Fn3 gene using

PCR. The expression of this phage results in the display of the fusion pIII

30   protein on the surface of the fd phage. Variegations in the Fn3 loops are

introduced using site-directed mutagenesis as described hereinabove, or by
subcloning the Fn3 libraries constructed in M13 phage into the fUSE5 vector.

EXAMPLE VIII

Other phage display libraries

5   T7 phage libraries (Novagen, Madison, WI) and bacterial pili expression
systems (Invitrogen) are also useful to express the Fn3 gene.

EXAMPLE IX

Isolation of polypeptides which bind to macromolecular structures

The selection of phage-displayed monobodies was performed following
10   the protocols of Barbas and coworkers (Rosenblum & Barbas, 1995). Briefly,
approximately 1 μg of a target molecule ("antigen") in sodium carbonate buffer
(100 mM, pH 8.5) was immobilized in the wells of a microtiter plate (Maxisorp,
Nunc) by incubating overnight at 4°C in an air tight container. After the
removal of this solution, the wells were then blocked with a 3% solution of BSA
15   (Sigma, Fraction V) in TBS by incubating the plate at 37°C for 1 hour. A
phagemid library solution (50 μl) containing approximately 10^{12} colony forming
units (cfu) of phagemid was absorbed in each well at 37°C for 1 hour. The wells
were then washed with an appropriate buffer (typically TBST, 50 mM Tris-HC1
(pH 7.5), 150 mM NaC1, and 0.5% Tween20) three times (once for the first
20   round). Bound phage were eluted by an acidic solution (typically, 0.1 M
glycine-HCl, pH 2.2; 50 μl) and recovered phage were immediately neutralized
with 3 μl of Tris solution. Alternatively, bound phage were eluted by incubating
the wells with 50 μl of TBS containing the antigen (1 - 10 μM). Recovered
phage were amplified using the standard protocol employing the XL1Blue cells
25   as the host (Sambrook, *et al.*). The selection process was repeated 5-6 times to
concentrate positive clones. After the final round, individual clones were picked
and their binding affinities and DNA sequences were determined.

The binding affinities of monobodies on the phage surface were
characterized using the phage ELISA technique (Li, *et al.*, 1995). Wells of
30   microtiter plates (Nunc) were coated with an antigen and blocked with BSA.
Purified phages (10^8- 10^{11} cfu) originating from a single colony were added to

each well and incubated 2 hours at 37°C. After washing wells with an

appropriate buffer (see above), bound phage were detected by the standard

ELISA protocol using anti-M13 antibody (rabbit, Sigma) and anti-rabbit Ig-

peroxidase conjugate (Pierce). Colorimetric assays were performed using

5   Turbo-TMB (3,3',5,5'-tetramethylbenzidine, Pierce) as a substrate.

The binding affinities of monobodies on the phage surface were further

characterized using the competition ELISA method (Djavadi-Ohaniance, et al.,

1996). In this experiment, phage ELISA is performed in the same manner as

described above, except that the phage solution contains a ligand at varied

10   concentrations. The phage solution was incubated a 4°C for one hour prior to

the binding of an immobilized ligand in a microtiter plate well. The affinities of

phage displayed monobodies are estimated by the decrease in ELISA signal as

the free ligand concentration is increased.

After preliminary characterization of monobodies displayed on the

15   surface of phage using phage ELISA, genes for positive clones were subcloned

into the expression vector pAS45. *E. coli* BL21(DE3) (Novagen) was

transformed with an expression vector (pAS45 and its derivatives). Cells were

grown in M9 minimal medium and M9 medium supplemented with

Bactotryptone (Difco) containing ampicillin (200 µg/ml). For isotopic labeling,

20   ^{15}N NH$_4$Cl and/or ^{13}C glucose replaced unlabeled components. Stable isotopes

were purchased from Isotec and Cambridge Isotope Labs. 500 ml medium in a 2

l baffle flask was inoculated with 10 ml of overnight culture and agitated at

approximately 140 rpm at 37°C. IPTG was added at a final concentration of 1

mM to induce protein expression when OD(600 nm) reached approximately 1.0.

25   The cells were harvested by centrifugation 3 hours after the addition of IPTG and

kept frozen at -70°C until used.

Fn3 and monobodies with His•tag were purified as follows. Cells were

suspended in 5 ml/(g cell) of 50 mM Tris (pH 7.6) containing 1 mM

phenylmethylsulfonyl fluoride. HEL (Sigma, 3X crystallized) was added to a

30   final concentration of 0.5 mg/ml. After incubating the solution for 30 min at

37°C, it was sonicated so as to cause cell breakage three times for 30 seconds on

39

ice. Cell debris was removed by centrifugation at 15,000 rpm in an Sorval RC-
2B centrifuge using an SS-34 rotor. Concentrated sodium chloride is added to
the solution to a final concentration of 0.5 M. The solution was then applied to a
1 ml HisTrap™ chelating column (Pharmacia) preloaded with nickel chloride

5   (0.1 M, 1 ml) and equilibrated in the Tris buffer (50 mM, pH 8.0) containing 0.5
M sodium chloride. After washing the column with the buffer, the bound protein
was eluted with a Tris buffer (50 mM, pH 8.0) containing 0.5 M imidazole. The
His•tag portion was cleaved off, when required, by treating the fusion protein
with thrombin using the protocol supplied by Novagen (Madison, WI). Fn3 was

10   separated from the His•tag peptide and thrombin by a Resources®column
(Pharmacia) using a linear gradient of sodium chloride (0 - 0.5 M) in sodium
acetate buffer (20 mM, pH 5.0).

       Small amounts of soluble monobodies were prepared as follows. XL-1
Blue cells containing pAS38 derivatives (plasmids coding Fn3-pIII fusion

15   proteins) were grown in LB media at 37°C with vigorous shaking until OD(600
nm) reached approximately 1.0; IPTG was added to the culture to a final
concentration of 1 mM, and the cells were further grown overnight at 37°C.
Cells were removed from the medium by centrifugation, and the supernatant was
applied to a microtiter well coated with a ligand. Although XL-1 Blue cells

20   containing pAS38 and its derivatives express FN3-pIII fusion proteins, soluble
proteins are also produced due to the cleavage of the linker between the Fn3 and
pIII regions by proteolytic activities of E. coli (Rosenblum & Barbas, 1995).
Binding of a monobody to the ligand was examined by the standard ELISA
protocol using a custom antibody against Fn3 (purchased from Cocalico

25   Biologicals, Reamstown, PA). Soluble monobodies obtained from the
periplasmic fraction of E. coli cells using a standard osmotic shock method were
also used.

40

EXAMPLE X

Ubiquitin binding monobody

Ubiquitin is a small (76 residue) protein involved in the degradation pathway in eurkaryotes. It is a single domain globular protein. Yeast ubiquitin

5   was purchased from Sigma Chemical Company and was used without further purification.

Libraries 2 and 4, described in Example VI above, were used to select ubiquitin-binding monobodies. Ubiquitin (1 µg in 50 µl sodium bicarbonate buffer (100 mM, pH 8.5)) was immobilized in the wells of a microtiter plate,

10   followed by blocking with BSA (3% in TBS). Panning was performed as described above. In the first two rounds, 1 µg of ubiquitin was immobilized per well, and bound phage were elute with an acidic solution. From the third to the sixth rounds, 0.1 µg of ubiquitin was immobilized per well and the phage were eluted either with an acidic solution or with TBS containing 10 µM ubiquitin.

15   Binding of selected clones was tested first in the polyclonal mode, i.e., before isolating individual clones. Selected clones from all libraries showed significant binding to ubiquitin. These results are shown in Figure 9. The binding to the immobilized ubiquitin of the clones was inhibited almost completely by less than 30 µM soluble ubiquitin in the competition ELISA

20   experiments (see Fig. 10). The sequences of the BC and FG loops of ubiquitin-binding monobodies is shown in Table 3.

Table 3. Sequences of ubiquitin-binding monobodies

	Name	BC loop	FG loop	Occurrence (if more than one)
	211	CARRA	RWIPLAK	2
5	212	CWRRA	RWVGLAW	
	213	CKHRR	FADLWWR	
	214	CRRGR	RGFMWLS	
	215	CNWRR	RAYRYRW	
	411	SRLRR	PPWRV	9
10	422	ARWTL	RRWWW	
	424	GQRTF	RRWWA	

The 411 clone, which was the most enriched clone, was characterized

using phage ELISA. The 411 clone showed selective binding and inhibition of

15   binding in the presence of about 10 µM ubiquitin in solution (Fig. 11).

EXAMPLE XI

Methods for the immobilization of small molecules

Target molecules were immobilized in wells of a microtiter plate

(Maxisorp, Nunc) as described hereinbelow, and the wells were blocked with

20   BSA. In addition to the use of carrier protein as described below, a conjugate of

a target molecule in biotin can be made. The biotinylated ligand can then be

immobilized to a microtiter plate well which has been coated with streptavidin.

In addition to the use of a carrier protein as described below, one could

make a conjugate of a target molecule and biotin (Pierce) and immobilize a

25   biotinylated ligand to a microtiter plate well which has been coated with

streptavidin (Smith and Scott, 1993).

Small molecules may be conjugated with a carrier protein such as bovine

serum albumin (BSA, Sigma), and passively adsorbed to the microtiter plate

well. Alternatively, methods of chemical conjugation can also be used. In

30   addition, solid supports other than microtiter plates can readily be employed.

42

EXAMPLE XII

Fluorescein binding monobody

Fluorescein has been used as a target for the selection of antibodies from combinatorial libraries (Barbas, *et al.* 1992). NHS-fluorescein was obtained
5   from Pierce and used according to the manufacturer's instructions in preparing conjugates with BSA (Sigma). Two types of fluorescein-BSA conjugates were prepared with approximate molar ratios of 17 (fluorescein) to one (BSA).

The selection process was repeated 5-6 times to concentrate positive clones. In this experiment, the phage library was incubated with a protein
10   mixture (BSA, cytochrome C (Sigma, Horse) and RNaseA (Sigma, Bovine), 1 mg/ml each) at room temperature for 30 minutes, prior to the addition to ligand coated wells. Bound phage were eluted in TBS containing 10 μM soluble fluorescein, instead of acid elution. After the final round, individual clones were picked and their binding affinities (see below) and DNA sequences were
15   determined.

Table 4. Clones from Library #2

		BC	FG
	WT	AVTVR	RGDSPAS
20			
	pLB24.1	CNWRR	RAYRYRW
	pLB24.2	CMWRA	RWGMLRR
	pLB24.3	ARMRE	RWLRGRY
	pLB24.4	CARRR	RRAGWGW
25	pLB24.5	CNWRR	RAYRYRW
	pLB24.6	RWRER	RHPWTER
	pLB24.7	CNWRR	RAYRYRW
	pLB24.8	ERRVP	RLLLWQR
	pLB24.9	GRGAG	FGSFERR
30	pLB24.11	CRWTR	RRWFDGA
	pLB24.12	CNWRR	RAYRYRW

43

Clones from Library #4

	WT	AVTVR	GRGDS
5	pLB25.1	GQRTF	RRWWA
	pLB25.2	GQRTF	RRWWA
	pLB25.3	GQRTF	RRWWA
	pLB25.4	LRYRS	GWRWR
	pLB25.5	GQRTF	RRWWA
10	pLB25.6	GQRTF	RRWWA
	pLB25.7	LRYRS	GWRWR
	pLB25.9	LRYRS	GWRWR
	pLB25.11	GQRTF	RRWWA
	pLB25.12	LRYRS	GWRWR

15

Preliminary characterization of the binding affinities of selected clones
were performed using phage ELISA and competition phage ELISA (see Fig. 12
(Fluorescein-1) and Fig. 13 (Fluorescein-2)). The four clones tested showed
20　specific binding to the ligand-coated wells, and the binding reactions are
inhibited by soluble fluorescein (see Fig. 13).

EXAMPLE XIII

Digoxigenin binding monobody

Digoxigenin-3-O-methyl-carbonyl-e-aminocapronic acid-NHS
25　(Boehringer Mannheim) is used to prepare a digoxigenin-BSA conjugate. The
coupling reaction is performed following the manufacturers' instructions. The
digoxigenin-BSA conjugate is immobilized in the wells of a microtiter plate and
used for panning. Panning is repeated 5 to 6 times to enrich binding clones.
Because digoxigenin is sparingly soluble in aqueous solution, bound phages are
30　eluted from the well using acidic solution. See Example XIV.

44

EXAMPLE XIV

TSAC (transition state analog compound) binding monobodies

Carbonate hydrolyzing monobodies are selected as follows. A transition

state analog for carbonate hydrolysis, 4-nitrophenyl phosphonate is synthesized

5   by an Arbuzov reaction as described previously (Jacobs and Schultz, 1987). The

phosphonate is then coupled to the carrier protein, BSA, using carbodiimide,

followed by exhaustive dialysis (Jacobs and Schultz, 1987). The hapten-BSA

conjugate is immobilized in the wells of a microtiter plate and monobody

selection is performed as described above. Catalytic activities of selected

10   monobodies are tested using 4-nitrophenyl carbonate as the substrate.

Other haptens useful to produce catalytic monobodies are summarized in

H. Suzuki (1994) and in N. R. Thomas (1994).

EXAMPLE XV

NMR characterization of Fn3 and comparison of the Fn3

15   secreted by yeast with that secreted by *E. coli*

Nuclear magnetic resonance (NMR) experiments are performed to

identify the contact surface between FnAb and a target molecule, e.g.,

monobodies to fluorescein, ubiquitin, RNaseA and soluble derivatives of

digoxigenin. The information is then be used to improve the affinity and

20   specificity of the monobody. Purified monobody samples are dissolved in an

appropriate buffer for NMR spectroscopy using Amicon ultrafiltration cell with a

YM-3 membrane. Buffers are made with 90 % H_2O/10 % D_2O (distilled grade,

Isotec) or with 100 % D_2O. Deuterated compounds (e.g. acetate) are used to

eliminate strong signals from them.

25   NMR experiments are performed on a Varian Unity INOVA 600

spectrometer equipped with four RF channels and a triple resonance probe with

pulsed field gradient capability. NMR spectra are analyzed using processing

programs such as Felix (Molecular Simulations), nmrPipe, PIPP, and CAPP

(Garrett, *et al.*, 1991; Delaglio, *et al.*, 1995) on UNIX workstations. Sequence

30   specific resonance assignments are made using well-established strategy using a

45

set of triple resonance experiments (CBCA(CO)NH and HNCACB) (Grzesiek &
Bax, 1992; Wittenkind & Mueller, 1993).

Nuclear Overhauser effect (NOE) is observed between ^{1}H nuclei closer
than approximately 5 Å, which allows one to obtain information on interproton

5   distances. A series of double- and triple-resonance experiments (Table 5; for
recent reviews on these techniques, see Bax & Grzesiek, 1993 and Kay, 1995)
are performed to collect distance (i.e. NOE) and dihedral angle (J-coupling)
constraints. Isotope-filtered experiments are performed to determine resonance
assignments of the bound ligand and to obtain distance constraints within the

10   ligand and those between FnAb and the ligand. Details of sequence specific
resonance assignments and NOE peak assignments have been described in detail
elsewhere (Clore & Gronenborn, 1991; Pascal, *et al.*, 1994b; Metzler, *et al.*,
1996).

15   Table 5. NMR experiments for structure characterization

Experiment Name	Reference
1. reference spectra	

2D-^{1}H, ^{15}N-HSQC	(Bodenhausen & Ruben, 1980; Kay, *et al.*, 1992)
2D-^{1}H, ^{13}C-HSQC	(Bodenhausen & Ruben, 1980; Vuister & Bax, 1992)

2. backbone and side chain resonance assignments of ^{13}C/^{15}N-labeled protein

3D-CBCA(CO)NH	(Grzesiek & Bax, 1992)
3D-HNCACB	(Wittenkind & Mueller, 1993)
3D-C(CO)NH	(Logan *et al.*, 1992; Grzesiek *et al.*, 1993)
3D-H(CCO)NH	
3D-HBHA(CBCACO)NH	(Grzesiek & Bax, 1993)
3D-HCCH-TOCSY	(Kay *et al.*, 1993)
3D-HCCH-COSY	(Ikura *et al.*, 1991)
3D-^{1}H, ^{15}N-TOCSY-HSQC	(Zhang *et al.*, 1994)
2D-HB(CBCDCE)HE	(Yamazaki *et al.*, 1993)

3. resonance assignments of unlabeled ligand

2D-isotope-filtered ^{1}H-TOCSY

5   2D-isotope-filtered ^{1}H-COSY

2D-isotope-filtered ^{1}H-NOESY   (Ikura & Bax, 1992)

4. structural constraints

within labeled protein

10   3D-^{1}H, ^{15}N-NOESY-HSQC   (Zhang *et al.*, 1994)

4D-^{1}H, ^{13}C-HMQC-NOESY-HMQC   (Vuister *et al.*, 1993)

4D-^{1}H, ^{13}C, ^{15}N-HSQC-NOESY-HSQC (Muhandiram *et al.*, 1993; Pascal *et al.*, 1994a)

within unlabeled ligand

2D-isotope-filtered ^{1}H-NOESY   (Ikura & Bax, 1992)

15   *interactions between protein and ligand*

3D-isotope-filtered ^{1}H, ^{15}N-NOESY-HSQC

3D-isotope-filtered ^{1}H, ^{13}C-NOESY-HSQC   (Lee *et al.*, 1994)

5. dihedral angle constraints

20

J-molulated ^{1}H, ^{15}N-HSQC   (Billeter *et al.*, 1992)

3D-HNHB   (Archer *et al.*, 1991)

      Backbone ^{1}H, ^{15}N and ^{13}C resonance assignments for a monobody are

compared to those for wild-type Fn3 to assess structural changes in the mutant.

25   Once these data establish that the mutant retains the global structure, structural

refinement is performed using experimental NOE data. Because the structural

difference of a monobody is expected to be minor, the wild-type structure can be

used as the initial model after modifying the amino acid sequence. The

mutations are introduced to the wild-type structure by interactive molecular

30   modeling, and then the structure is energy-minimized using a molecular

modeling program such as Quanta (Molecular Simulations). Solution structure

is refined using cycles of dynamical simulated annealing (Nilges *et al.*, 1988) in

the program X-PLOR (Brünger, 1992). Typically, an ensemble of fifty

47

structures is calculated. The validity of the refined structures is confirmed by calculating a fewer number of structures from randomly generated initial structures in X-PLOR using the YASAP protocol (Nilges, et al., 1991). Structure of a monobody-ligand complex is calculated by first refining both
5   components individually using intramolecular NOEs, and then docking the two using intermolecular NOEs.

For example, the ^{1}H, ^{15}N-HSQC spectrum for the fluorescein-binding monobody LB25.5 is shown in Figure 14. The spectrum shows a good dispersion (peaks are spread out) indicating that LB25.5 is folded into a globular
10   conformation. Further, the spectrum resembles that for the wild-type Fn3, showing that the overall structure of LB25.5 is similar to that of Fn3. These results demonstrate that ligand-binding monobodies can be obtained without changing the global fold of the Fn3 scaffold.

Chemical shift perturbation experiments are performed by forming the
15   complex between an isotope-labeled FnAb and an unlabeled ligand. The formation of a stoichiometric complex is followed by recording the HSQC spectrum. Because chemical shift is extremely sensitive to nuclear environment, formation of a complex usually results in substantial chemical shift changes for resonances of amino acid residues in the interface. Isotope-edited NMR
20   experiments (2D HSQC and 3D CBCA(CO)NH) are used to identify the resonances that are perturbed in the labeled component of the complex; i.e. the monobody. Although the possibility of artifacts due to long-range conformational changes must always be considered, substantial differences for residues clustered on continuous surfaces are most likely to arise from direct
25   contacts (Chen et al., 1993; Gronenborn & Clore, 1993).

An alternative method for mapping the interaction surface utilizes amide hydrogen exchange (HX) measurements. HX rates for each amide proton are measured for ^{15}N labeled monobody both free and complexed with a ligand. Ligand binding is expected to result in decreased amide HX rates for monobody
30   residues in the interface between the two proteins, thus identifying the binding surface. HX rates for monobodies in the complex are measured by allowing HX

48

to occur for a variable time following transfer of the complex to D_2O; the

complex is dissociated by lowering pH and the HSQC spectrum is recorded at

low pH where amide HX is slow. Fn3 is stable and soluble at low pH, satisfying

the prerequisite for the experiments.

5   **EXAMPLE XVI**

Construction and Analysis of Fn3-Display System Specific for Ubiquitin

An Fn3-display system was designed and synthesized, ubiquitin-binding

clones were isolated and a major Fn3 mutant in these clones was biophysically

characterized.

10   Gene construction and phage display of Fn3 was performed as in

Examples I and II above. The Fn3-phage pIII fusion protein was expressed from

a phagemid-display vector, while the other components of the M13 phage,

including the wild-type pIII, were produced using a helper phage (Bass et al.,

1990). Thus, a phage produced by this system should contain less than one copy

15   of Fn3 displayed on the surface. The surface display of Fn3 on the phage was

detected by ELISA using an anti-Fn3 antibody. Only phages containing the Fn3-

pIII fusion vector reacted with the antibody.

After confirming the phage surface to display Fn3, a phage display

library of Fn3 was constructed as in Example III. Random sequences were

20   introduced in the BC and FG loops. In the first library, five residues (77-81)

were randomized and three residues (82-84) were deleted from the FG loop. The

deletion was intended to reduce the flexibility and improve the binding affinity

of the FG loop. Five residues (26-30) were also randomized in the BC loop in

order to provide a larger contact surface with the target molecule. Thus, the

25   resulting library contains five randomized residues in each of the BC and FG

loops (Table 6). This library contained approximately 10^8 independent clones.

Library Screening

Library screening was performed using ubiquitin as the target molecule.

In each round of panning, Fn3-phages were absorbed to a ubiquitin-coated

30   surface, and bound phages were eluted competitively with soluble ubiquitin.

The recovery ratio improved from 4.3×10^{-7} in the second round to 4.5×10^{-6} in

the fifth round, suggesting an enrichment of binding clones. After five founds of

panning, the amino acid sequences of individual clones were determined (Table

6).

Table 6. Sequences in the variegated loops of enriched clones

Name	BC loop	FG loop	Frequency
Wild Type	GCAGTTACCGTGCGT	GGCCGTGGTGACAGCCCAGCGAGC	–
	AlaValThrValArg	GlyArgGlyAspSerProAlaSer	
Library*	NNKNNKNNKNNKNNK	NNKNNKNNKNNKNNK--------	–
	X X X X X	X X X X X (deletion)	
clone1	TCGAGGTTGCGGCGG	CCGCCGTGGAGGGTG	9
(Ubi4)	SerArgLeuArgArg	ProProTrpArgVal	
clone2	GGTCAGCGAACTTTT	AGGCGGTGGTGGGCT	1
	GlyGlnArgThrPhe	ArgArgTrpTrpAla	
clone3	GCGAGGTGGACGCTT	AGGCGGTGGTGGTGG	1
	AlaArgTrpThrLeu	ArgArgTrpTrpTrp	

* N denotes an equimolar mixture of A, T, G and C; K denotes an equimolar mixture of G and T.

A clone, dubbed Ubi4, dominated the enriched pool of Fn3 variants. Therefore,

further investigation was focused on this Ubi4 clone. Ubi4 contains four

mutations in the BC loop (Arg 30 in the BC loop was conserved) and five

mutations and three deletions in the FG loop. Thus 13% (12 out of 94) of the

residues were altered in Ubi4 from the wild-type sequence.

Figure 15 shows a phage ELISA analysis of Ubi4. The Ubi4 phage binds

to the target molecule, ubiquitin, with a significant affinity, while a phage

displaying the wild-type Fn3 domain or a phase with no displayed molecules

show little detectable binding to ubiquitin (Figure 15a). In addition, the Ubi4

phage showed a somewhat elevated level of background binding to the control

surface lacking the ubiquitin coating. A competition ELISA experiments shows

the IC_{50} (concentration of the free ligand which causes 50% inhibition of

binding) of the binding reaction is approximately 5 μM (Fig. 15b). BSA, bovine

ribonuclease A and cytochrome C show little inhibition of the Ubi4-ubiquitin

binding reaction (Figure 15c), indicating that the binding reaction of Ubi4 to

ubiquitin does result from specific binding.

50

Characterization of a Mutant Fn3 Protein

The expression system yielded 50-100 mg Fn3 protein per liter culture.
A similar level of protein expression was observed for the Ubi4 clone and other
mutant Fn3 proteins.

5   Ubi4-Fn3 was expressed as an independent protein. Though a majority
of Ubi4 was expressed in E. coli as a soluble protein, its solubility was found to
be significantly reduced as compared to that of wild-type Fn3. Ubi4 was soluble
up to ~20 µM at low pH, with much lower solubility at neutral pH. This
solubility was not high enough for detailed structural characterization using

10   NMR spectroscopy or X-ray crystallography.

The solubility of the Ubi4 protein was improved by adding a solubility
tail, GKKGK, as a C-terminal extension. The gene for Ubi4-Fn3 was subcloned
into the expression vector pAS45 using PCR. The C-terminal solubilization tag,
GKKGK, was incorporated in this step. E. coli BL21 (DE3) (Novagen) was

15   transformed with the expression vector (pAS45 and its derivatives). Cells were
grown in M9 minimal media and M9 media supplemented with Bactotryptone
(Difco) containing ampicillin (200 µg/ml). For isotopic labeling, ^{15}N NH$_4$Cl
replaced unlabeled NH$_4$Cl in the media. 500 ml medium in a 2 liter baffle flask
was inoculated with 10 ml of overnight culture and agitated at 37°C. IPTG was

20   added at a final concentration of 1 mM to initiate protein expression when OD
(600 nm) reaches one. The cells were harvested by centrifugation 3 hours after
the addition of IPTG and kept frozen at -70°C until used.

Proteins were purified as follows. Cells were suspended in 5 ml/(g cell)
of Tris (50 mM, pH 7.6) containing phenylmethylsulfonyl fluoride (1 mM). Hen

25   egg lysozyme (Sigma) was added to a final concentration of 0.5 mg/ml. After
incubating the solution for 30 minutes at 37°C, it was sonicated three times for
30 seconds on ice. Cell debris was removed by centrifugation. Concentrated
sodium chloride was added to the solution to a final concentration of 0.5 M. The
solution was applied to a Hi-Trap chelating column (Pharmacia) preloaded with

30   nickel and equilibrated in the Tris buffer containing sodium chloride (0.5 M).
After washing the column with the buffer, histag-Fn3 was eluted with the buffer

containing 500 mM imidazole. The protein was further purified using a
ResourceS column (Pharmacia) with a NaCl gradient in a sodium acetate buffer
(20 mM, pH 4.6).

With the GKKGK tail, the solubility of the Ubi4 protein was increased to

5   over 1 mM at low pH and up to ~50 µM at neutral pH. Therefore, further
analyses were performed on Ubi4 with this C-terminal extension (hereafter
referred to as Ubi4-K). It has been reported that the solubility of a minibody
could be significantly improved by addition of three Lys residues at the N- or C-
termini (Bianchi et al., 1994). In the case of protein Rop, a non-structured C-

10   terminal tail is critical in maintaining its solubility (Smith et al., 1995).

Oligomerization states of the Ubi4 protein were determined using a size
exclusion column. The wild-type Fn3 protein was monomeric at low and neutral
pH's. However, the peak of the Ubi4-K protein was significantly broader than
that of wild-type Fn3, and eluted after the wild-type protein. This suggests

15   interactions between Ubi4-K and the column material, precluding the use of size
exclusion chromatography to determine the oligomerization state of Ubi4. NMR
studies suggest that the protein is monomeric at low pH.

The Ubi4-K protein retained a binding affinity to ubiquitin as judged by
ELISA (Figure 15d). However, an attempt to determine the dissociation

20   constant using a biosensor (Affinity Sensors, Cambridge, U.K.) failed because of
high background binding of Ubi4-K-Fn3 to the sensor matrix. This matrix
mainly consists of dextran, consistent with our observation that interactions
between Ubi4-K interacts with the cross-linked dextran of the size exclusion
column.

25   **Example XVII**

                  Stability Measurements of Monobodies

Guanidine hydrochloride (GuHCl)-induced unfolding and refolding
reactions were followed by measuring tryptophan fluorescence. Experiments
were performed on a Spectronic AB-2 spectrofluorometer equipped with a

30   motor-driven syringe (Hamilton Co.). The cuvette temperature was kept at
30°C. The spectrofluorometer and the syringe were controlled by a single

52

computer using a home-built interface. This system automatically records a
series of spectra following GuHCl titration. An experiment started with a 1.5 ml
buffer solution containing 5 μM protein. An emission spectrum (300-400 nm;
excitation at 290 nm) was recorded following a delay (3-5 minutes) after each

5   injection (50 or 100 μl) of a buffer solution containing GuHCl. These steps were
repeated until the solution volume reached the full capacity of a cuvette (3.0 ml).
Fluorescence intensities were normalized as ratios to the intensity at an
isofluorescent point which was determined in separate experiments. Unfolding
curves were fitted with a two-state model using a nonlinear least-squares routine

10   (Santoro & Bolen, 1988). No significant differences were observed between
experiments with delay times (between an injection and the start of spectrum
acquisition) of 2 minutes and 10 minutes, indicating that the unfolding/refolding
reactions reached close to an equilibrium at each concentration point within the
delay times used.

15   Conformational stability of Ubi4-K was measured using above-described
GuHCl-induced unfolding method. The measurements were performed under
two sets of conditions; first at pH 3.3 in the presence of 300 mM sodium
chloride, where Ubi4-K is highly soluble, and second in TBS, which was used
for library screening. Under both conditions, the unfolding reaction was

20   reversible, and we detected no signs of aggregation or irreversible unfolding.
Figure 16 shows unfolding transitions of Ubi4-K and wild-type Fn3 with the N-
terminal (his)$_6$ tag and the C-terminal solubility tag. The stability of wild-type
Fn3 was not significantly affected by the addition of these tags. Parameters
characterizing the unfolding transitions are listed in Table 7.

25

Table 7. Stability parameters for Ubi4 and wild-type Fn3 as determined by GuHCl-induced unfolding

Protein	ΔG_0 (kcal mol^{-1})	m_G (kcal mol^{-1} M^{-1})
Ubi4 (pH 7.5)	4.8 ± 0.1	2.12 ± 0.04
Ubi4 (pH 3.3)	6.5 ± 0.1	2.07 ± 0.02
Wild-type (pH 7.5)	7.2 ± 0.2	1.60 ± 0.04
Wild-type (pH 3.3)	11.2 ± 0.1	2.03 ± 0.02

ΔG_0 is the free energy of unfolding in the absence of denaturant; m_G is the dependence of the free energy of unfolding on GuHCl concentration. For solution conditions, see Figure 4 caption.

Though the introduced mutations in the two loops certainly decreased the stability of Ubi4-K relative to wild-type Fn3, the stability of Ubi4 remains comparable to that of a "typical" globular protein. It should also be noted that the stabilities of the wild-type and Ubi4-K proteins were higher at pH 3.3 than at pH 7.5.

The Ubi4 protein had a significantly reduced solubility as compared to that of wild-type Fn3, but the solubility was improved by the addition of a solubility tail. Since the two mutated loops comprise the only differences between the wild-type and Ubi4 proteins, these loops must be the origin of the reduced solubility. At this point, it is not clear whether the aggregation of Ubi4-K is caused by interactions between the loops, or by interactions between the loops and the invariable regions of the Fn3 scaffold.

The Ubi4-K protein retained the global fold of Fn3, showing that this scaffold can accommodate a large number of mutations in the two loops tested. Though the stability of the Ubi4-K protein is significantly lower than that of the wild-type Fn3 protein, the Ubi4 protein still has a conformational stability comparable to those for small globular proteins. The use of a highly stable domain as a scaffold is clearly advantageous for introducing mutations without affecting the global fold of the scaffold. In addition, the GuHCl-induced unfolding of the Ubi4 protein is almost completely reversible. This allows the preparation of a correctly folded protein even when a Fn3 mutant is expressed in

54

a misfolded form, as in inclusion bodies. The modest stability of Ubi4 in the conditions used for library screening indicates that Fn3 variants are folded on the phage surface. This suggests that a Fn3 clone is selected by its binding affinity in the folded form, not in a denatured form. Dickinson et al. proposed that Val

5   29 and Arg 30 in the BC loop stabilize Fn3. Val 29 makes contact with the hydrophobic core, and Arg 30 forms hydrogen bonds with Gly 52 and Val 75. In Ubi4-Fn3, Val 29 is replaced with Arg, while Arg 30 is conserved. The FG loop was also mutated in the library. This loop is flexible in the wild-type structure, and shows a large variation in length among human Fn3 domains (Main et al.,

10   1992). These observations suggest that mutations in the FG loop may have less impact on stability. In addition, the N-terminal tail of Fn3 is adjacent to the molecular surface formed by the BC and FG loops (Figure 1 and 17) and does not form a well-defined structure. Mutations in the N-terminal tail would not be expected to have strong detrimental effects on stability. Thus, residues in the N-

15   terminal tail may be good sites for introducing additional mutations.

Example XVIII

NMR Spectroscopy of Ubi4-Fn3

Ubi4-Fn3 was dissolved in [^{2}H]-Gly HCl buffer (20 mM, pH 3.3) containing NaCl (300 mM) using an Amicon ultrafiltration unit. The final

20   protein concentration was 1 mM. NMR experiments were performed on a Varian Unity INOVA 600 spectrometer equipped with a triple-resonance probe with pulsed field gradient. The probe temperature was set at 30°C. HSQC, TOCSY-HSQC and NOESY-HSQC spectra were recorded using published procedures (Kay et al., 1992; Zhang et al., 1994). NMR spectra were processed

25   and analyzed using the NMRPipe and NMRView software (Johnson & Blevins, 1994; Delaglio et al., 1995) on UNIX workstations. Sequence-specific resonance assignments were made using standard procedures (Wüthrich, 1986; Clore & Gronenborn, 1991). The assignments for wild-type Fn3 (Baron et al., 1992) were confirmed using a ^{15}N-labeled protein dissolved in sodium acetate

30   buffer (50 mM, pH 4.6) at 30°C.

55

The three-dimensional structure of Ubi4-K was characterized using this heteronuclear NMR spectroscopy method. A high quality spectrum could be collected on a 1 mM solution of ^{15}N-labeled Ubi4 (Figure 17a) at low pH. The linewidth of amide peaks of Ubi4-K was similar to that of wild-type Fn3,

5   suggesting that Ubi4-K is monomeric under the conditions used. Complete assignments for backbone ^{1}H and ^{15}N nuclei were achieved using standard ^{1}H, ^{15}N double resonance techniques, except for a row of His residues in the N-terminal (His)$_6$ tag. There were a few weak peaks in the HSQC spectrum which appeared to originate from a minor species containing the N-terminal Met

10   residue. Mass spectroscopy analysis showed that a majority of Ubi4-K does not contain the N-terminal Met residue. Fig. 17 shows differences in ^{1}HN and ^{15}N chemical shifts between Ubi4-K and wild-type Fn3. Only small differences are observed in the chemical shifts, except for those in and near the mutated BC and FG loops. These results clearly indicate that Ubi4-K retains the global fold of

15   Fn3, despite the extensive mutations in the two loops. A few residues in the N-terminal region, which is close to the two mutated loops, also exhibit significant chemical differences between the two proteins. An HSQC spectrum was also recorded on a 50 μM sample of Ubi4-K in TBS. The spectrum was similar to that collected at low pH, indicating that the global conformation of Ubi4 is

20   maintained between pH 7.5 and 3.3.

The foregoing detailed description and examples have been given for clarity of understanding only. No unnecessary limitations are to be understood therefrom. The invention is not limited to the exact details shown and described for variations obvious to one skilled in the art will be included within the

25   invention defined by the claims.

56

REFERENCES

Alzari, P.N., Lascombe, M.-B. & Poljak, R.J. (1988) Three-dimensional structure of antibodies. *Annu. Rev. Immunol. 6*, 555-580.

Archer, S. J., Ikura, M., Torchia, D. A. & Bax, A. (1991) An alternative 3D NMR technique for correlating backbone 15N with side chain Hb resonances in large proteins *J. Magn. Reson. 95*, 636-641.

Aukhil, I., Joshi, P., Yan, Y. & Erickson, H. P. (1993) Cell- and heparin-binding domains of the hexabrachion arm identified by tenascin expression protein *J. Biol. Chem. 268*, 2542-2553.

Barbas, C. F., III, Kang, A. S., Lerner, R. A. & Benkovic, S. J. (1991) Assembly of combinatorial antibody libraries on phage surfaces: the gene III site. *Proc. Natl. Acad. Sci. USA 88*, 7978-7982.

Barbas, C. F., III, Bain, J. D., Hoekstra, D. M. & Lerner, R. A. (1992) Semisynthetic combinatorial libraries: A chemical solution to the diversity problem *Proc. Natl. Acad. Sci. USA 89*, 4457-4461.

Baron, M., Main, A. L., Driscoll, P. C., Mardon, H. J., Boyd, J. & Campbell, I.D. (1992) ^{1}H NMR assignment and secondary structure of the cell adhesion type II module of fibronectin *Biochemistry 31*, 2068-2073.

Baron, M., Norman, D. G. & Campbell, I. D. (1991) Protein modules *Trends Biochem. Sci. 16*, 13-17.

Bass, S., Greene, R. & Wells, J. A. (1990) Hormone phage: An enrichment method for variant proteins with altered binding properties *Proteins: Struct. Funct. Genet. 8*, 309-314.

Bax, A. & Grzesiek, S. (1993) Methodological advances in protein NMR. *Acc. Chem. Res. 26*, 131-138.

Becktel, W. J. & Schellman, J. A. (1987) Protein stability curves. *Biopolymer 26*, 1859-1877.

Bhat, T. N., Bentley, G. A., Boulot, G., Greene, M. I., Tello, D., Dall'acqua, W., Souchon, H., Schwarz, F. P., Mariuzza, R. A. & Poljak, R. J. (1994) Bound water molecules and conformational stabilization help mediate an antigen-antibody association. *Proc. Natl. Acad. Sci. USA 91*, 1089-1093.

Bianchi, E., Venturini, S., Pessi, A., Tramontano, A. & Sollazzo, M. (1994)
High level expression and rational mutagenesis of a designed protein, the
minibody. From an insoluble to a soluble molecule. *J. Mol. Biol. 236*,
649.659.

Billeter, M., Neri, D., Otting, G., Qian, Y. Q. & Wüthrich, K. (1992) Precise
vicinal coupling constants 3JHNa in proteins from nonlinear fits of J-
modulated [^{15}N, ^{1}H]-COSY experiments. *J. Biomol. NMR 2*, 257-274.

Bodenhausen, G. & Ruben, D. J. (1980) Natural abundance nitrogen-15 NMR by
enhanced heteronuclear spectroscopy. *Chem. Phys. Lett. 69*, 185-189.

Bork, P. & Doolittle, R. F., *PNAS* 89:8990-8994 (1992).

Bork, P. & Doolittle, R. F. (1992) Proposed acquisition of an animal protein
domain by bacteria. *Proc. Natl. Acad. Sci. USA 89*, 8990-8994.

Bork, P., Hom, L. & Sander, C. (1994) The immunoglobulin fold. Structural
classification, sequence patterns and common core. *J. Mol. Biol. 242*,
309-320.

Brünger, A. T. (1992) *X-PLOR (Version 3.1): A system for X-ray
crystallography and NMR.*, Yale Univ. Press, New Haven.

Burke, T., Bolger, R., Checovich, W. & Lowery, R. (1996) in *Phage display of
peptides and proteins* (Kay, B. K., Winter, J. and McCafferty, J., Ed.)
Vol. pp305-326, Academic Press, San Diego.

Campbell, I. D. & Spitzfaden, C. (1994) Building proteins with fibronectin type
III modules *Structure 2*, 233-337.

Chen, Y., Reizer, J., Saier, M. H., Fairbrother, W. J. & Wright, P. E. (1993)
Mapping the binding interfaces of the proteins of the bacterial
phaphotransferase system, HPr and IIAglc. *Biochemistry 32*, 32-37.

Clackson & Wells, (1994) *Trends Biotechnology 12*, 173-184.

Clore, G. M. & Gronenborn, A. M. (1991) Structure of larger proteins in
solution: Three- and four-dimensional heteronuclear NMR spectroscopy.
Science 252, 1390-1399.

Corey, D. R., Shiau, A. K., Q., Y., Janowski, B. A. & Craik, C. S. (1993)
Trypsin display on the surface of bacteriophage. *Gene 128*, 129-134.

58

Davies, J. & Riechmann, L. (1996). Single antibody domains as small
    recognition units: design and *in vitro* antigen selection of camelized,
    human VH domains with improved protein stability. Protein Eng., 9(6),
    531-537.

Davies, J. & Riechmann, L. (1995) Antobody VH domains as small recognition
    units. *Bio/Technol. 13*, 475-479.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995)
    NMRPipe: a multidimensional spectral processing system based on
    UNIX pipes. *J. Biomol. NMR 6*, 277-293.

Deng, W. P. & Nickoloff, J. A. (1992) Site-directed mutagenesis of virtually any
    plasmid by eliminating a unique site. *Anal. Biochem. 200*, 81-88.

deVos, A. M., Ultsch, M. & Kossiakoff, A. A. (1992) Human Growth hormone
    and extracellular domain of its receptor: crystal structure of the complex.
    Science 255, 306-312.

Dickinson, C. D., Veerapandian, B., Dai, X.P., Hamlin, R. C., Xuong, N.-H.,
    Ruoslahti, E. & Ely, K. R. (1994) Crystal structure of the tenth type III
    cell adhesion module of human fibronectin *J. Mol. Biol. 236*, 1079-1092.

Djavadi-Ohaniance, L., Goldberg, M. E. & Friguet, B. (1996) in *Antibody
    Engineering. A Practical Approach* (McCafferty, J., Hoogenboom, H. R.
    and Chiswell, D. J., Ed.) Vol. pp77-97, Oxford Univ. Press, Oxford.

Dougall, W. C., Peterson, N. C. & Greene, M. I. (1994) Antibody-structure-
    based design of pharmacological agnets. *Trends Biotechnol. 12*, 372-
    379.

Garrett, D. S., Powers, R., Gronenborn, A. M. & Clore, G. M. (1991) A common
    sense approach to peak picking in two-, three- and four-dimensional
    spectra using automatic computer analysis of contour diagrams. *J. Magn.
    Reson. 95*, 214-220.

Ghosh, G., Van Duyne, G., Ghosh, S. & Sigler, P. B. (1995) Structure of NF-kB
    p50 homodimer bound to a kB site. *Nature 373*, 303-310.

59

Gribskov, M., Devereux, J. & Burgess, R. R. (1984) The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expression. *Nuc. Acids. Res. 12*, 539-549.

Groneborn, A. M., Filpula, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield, P. T. & Clore, G. M. (1991) A novel, highly stabel fold of the immunoglobulin binding domain of Streptococcal protein G. *Science 253*, 657-661.

Gronenborn, A. M. & Clore, G. M. (1993) Identification of the contact surface of a Streptococcal protein G domain complexed with a human Fc fragment. *J. Mol. Biol. 233*, 331-335.

Grzesiek, S., Anglister, J. & Bax, A. (1993) Correlation of backbone amide and aliphatic side-chain resonances in 13C/15N-enriched proteins by isotropic mixing of 13C magnetization. *J. Magn. Reson. B 101*, 114-119.

Grzesiek, S. & Bax, A. (1992) Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc. 114*, 6291-6293.

Grzesiek, S. & Bax, A. (1993) Amino acid type determination in the sequential assignment procedure of unformly 13C/15N-enriched proteins. *J. Biomol. NMR 3*, 185-204.

Harlow, E. & Lane, D. (1988) *Antibodies. A laboratory manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor.

Harpez, Y. & Chothia, C. (1994) Many of the immunoglobulin superfamily domains in cell adhesioin molecules and surface receptors belong to a new structural set which is close to that containing variable domains *J. Mol. Biol. 238*, 528-539.

Hawkins, R. E., Russell, S. J., Baier, M. & Winter, G. (1193) The contribution of contact and non-contact residues of antibody in the affinity of binding to antigen. The interaction of mutant D1.3 antibodies with lysozyme. *J. Mol. Biol. 234*, 958-964.

Holliger, P. et al., (1993) *Proc. Natl. Acad. Sci. 90*, 6444-6448.

Hu, S-z., et al., *Cancer Res.* 56:3055-3061 (1996).

Ikura, M. & Bax, A. (1992) Isotope-filtered 2D NMR of a protein-peptide
    complex: study of a skeletal muscle myosin light chain kinase fragment
    bound to calmodulin. *J. Am. Chem. Soc. 114*, 2433-2440.

Ikura, M., Kay, L. E. & Bax, A. (1991) Improved three-dimensional 1H-13C-1H
    correlation spectroscopy of a 13C-labeled protein using constant-time
    evolution. *J. Biomol. NMR 1*, 299-304.

Jacobs, J. & Schultz, P. G. (1987) Catalytic antibodies. *J. Am. Chem. Soc. 109*,
    2174-2176.

Janda, K. D., et al., *Science* 275:945-948 (1997).

Jones, E. Y. (1993) The immunoglobulin superfamily *Curr. Opinion struct. Biol.*
    3, 846-852.

Jones, P. T., Dear, P. H., Foote, J., Neuberger, M. S. & Winter, G. (1986)
    Replacing the complementarity-determining regions in a human antibody
    with those from a mouse *Nature 321*, 522-525.

Kabsch, W. & Sander, C. (1983) Dictionary of protein secondary structure:
    pattern recognition of hydrogen-bonded and geometrical features.
    Biopolymers 22, 2577-2637.

Kay, L. E. (1995) Field gradient techniques in NMR spectroscopy. *Curr.*
    Opinion Struct. Biol. 5, 674-681.

Kay, L. E., Keifer, P. & Saarinen, T. (1992) Pure absorption gradient enhanced
    heteronuclear single quantum correlation spectroscopy with improved
    sensitivity *J. Am. Chem. Soc. 114*, 10663-10665.

Kay, L. E., Xu, G.-Y. & Singer, A. U. (1993) A Gradient-Enhanced HCCH-
    TOCSY Experiment for Recording Side-Chain 1H and 13C Correlations
    in H2O Samples of Proteins. *J. Magn. Reson. B 101*, 333-337.

Koide, S., Dyson, H. J. & Wright, P. E. (1993) Characterization of a folding
    intermediate of apoplastcyanin trapped by proline isomerization.
    Biochemistry 32, 12299-12310.

Kornblihtt, A. R., Umezawa, K., Vibe-Pederson, K. & Baralle, F. E. (1985)
    Primary structure of human fibronectin: differential splicing may

61

generate at least 10 polypeptides from a single gene *EMBO J. 4*, 1755-1759.

Kraulis, P. (1991) MOLSCRIPT: a program to produce both detailed and
    scnematic plots of protein structures. *J. Appl. Cryst. 24*, 946-950.

Kunkel, T. A. (1985) Rapid and efficient site-specific mutagenesis without
    phenotypic selection. *Proc. Natl. Acad. Sci. USA 82*, 488-492.

Leahy, D. J., Hendrickson, W. A., Aukhil, I. & Erickson, H. P. (1992) Structure
    of a fibronectin type III domain from tenascin phased by MAD analysis
    of the selenomethionlyl protein *Science 258*, 987-991.

Lee, W., Revington, M. J., Arrowsmith, C. & Kay, L. E. (1994) A pulsed field
    gradient isotope-filtered 3D 13C HMQC-NOESY experiment for
    extracting intermolecular NOE contacts in molecular complexes. *FEBS
    lett. 350*, 87-90.

Lerner, R. A. & Barbas III, C. F., *Acta Chemica Scandinavica*, 50 672-678
    (1996).

Lesk, A. M. & Tramontano, A. (1992) Antibody structure and structural
    predictions useful in guiding antibody engineering. In *Antibody
    engineering. A practical guide.* (Borrebaeck, C. A. K., Ed.) Vol. W. H.
    Freeman & Co., New York.

Li, B., Tom, J. Y., Oare, D., Yen, R., Fairbrother, W. J., Wells, J. A. &
    Cunningham, B. C. (1995) Minimization of a polypeptide hormone
    Science 270, 1657-1660.

Litvinovich, S. V., Novokhatny, V. V., Brew, S. A. & Inhgam, K. C. (1992)
    Reversible unfolding of an isolated heparin and DNA binding fragment,
    the first type III module from fibronectin. *Biochim. Biophys. Acta 1119*,
    57-62.

Logan, T. M., Olejniczak, E. T., Xu, R. X. & Fesik, S. W. (1992) Side chain and
    backbone assignments in isotopically labeled proteins from two
    heteronuclear triple resonance experiments. *FEBS lett. 314*, 413-418.

Main, A. L., Harvey, T. S., Baron, M., Boyd, J. & Campbell, I. D. (1992) The
    three-dimensional structure of the tenth type III module of fibronectin: an
    insight into RGD-mediated interactions. *Cell 71*, 671-678.

62

Masat, L., et al., (1994) *PNAS* 91:893-896.

Martin, F., Toniatti, C., Ciliberto, G., Cortese, R. & Sollazzo, M. (1994) The
affinity-selection of a minibody polypeptide inhibitor of human
interleukin-6. *EMBO J. 13*, 5303-5309.

Martin, M. T., *Drug Discov. Today*, 1:239-247 (1996).

McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. (1990) Phage
antibodies: filamentous phage displaying antibody variable domains.
Nature 348, 552-554.

McConnell, S. J., & Hoess, R. H., *J.Mol. Biol.* 250:460-470 (1995).

Metzler, W. J., Leiting, B., Pryor, K., Mueller, L. & Farmer, B. T. I. (1996) The
three-dimensional solution structure of the SH2 domain from p55blk
kinase. *Biochemistry 35*, 6201-6211.

Minor, D. L. J. & Kim, P. S. (1994) Measurement of the β-sheet-forming
propensities of amino acids. *Nature 367*, 660-663.

Muhandiram, D. R., Xu, G. Y. & Kay, L. E. (1993) An enhanced-sensitivity pure
absorption gradient 4D 15N,13C-edited NOESY experiment. *J. Biomol.
NMR 3*, 463-470.

Müller, C. W., Rey, F. A., Sodeoka, M., Verdine, G. L. & Harrison, S. C. (1995)
Structure of the NH-kB p50 homodimer boound to DNA. *Nature 373*,
311-117.

Nilges, M., Clore, G. M. & Gronenborn, A. M. (1988) Determination of three-
dimensional structures of proteins from interproton distance data by
hybrid distance geometry-dynamical simulated annealing calculations.
FEBS lett. 229, 317-324.

Nilges, M., Kuszewski, J. & Brünger, A. T. (1991) in *Computational aspects of
the study of biological macromolecules by nuclear magnetic resonance
spectroscopy*. (Hoch, J. C., Poulsen, F. M. and Redfield, C., Ed.) Vol. pp.
451-455, Plenum Press, New York.

O'Neil et al., (1994) in *Techniques in Protein Chemistry V* (Crabb, L., ed.)
pp.517-524, Academic Press, San Diego.

63

O'Neil, K. T. & Hoess, R. H. (1995) Phage display: protein engineering by

    directed evolution. *Curr. Opinion Struct. Biol. 5*, 443-449.

Pace, C. N., Shirley, B. A. & Thomson, J. A. (1989) Measuring the

    conformational stability of a protein. In *Protein structure. A practical

    approach* (Creighton, T. E., Ed.) Vol. pp 311-330, IRL Press, Oxford.

Parmley, S. F. & Smith, G. P. (1988) Antibody-selectable filamentous fd phage

    vectors: affinity purification of target genes *Gene 73*, 305-318.

Pascal, S. M., Muhandiram, D. R., Yamazaki, T., Forman-Kay, J. D. & Kay, L.

    E. (1994a) Simultaneous acquisition of 15N- and 13C-edited NOE

    spectra of proteins dissolved in H2O. *J. Magn. Reson. B 103*, 197-201.

Pascal, S. M., Singer, A. U., Gish, G., Yamazaki, T., Shoelson, S. E., Pawson,

    T., Kay, L. E. & Forman-Kay, J. D. (1994b) Nuclear magnetic resonance

    structure of an SH2 domain of phospholipase C-g1 complexed with a high

    affinity binding peptide. *Cell 77*, 461-472.

Pessi, A., Bianchi, E., Crameri, A., Venturini, S., Tramontano, A. & Sollazzo,

    M. (1993) A designed metal-binding protein with a novel fold. *Nature

    362*, 3678-369.

Rees, A. R., Staunton, D., Webster, D. M., Searle, S. J., Henry, A. H. &

    Pedersen, J. T. (1994) Antibody design: beyond the natural limits. *Trends

    Biotechnol.* 12, 199-206.

Roberts et al., (1992) *Proc. Natl. Acad. Sci. USA 89*, 2429-2433.

Rosenblum, J. S. & Barbas, C. F. I. (1995) in *Antobody Engineering* (Borrenbaeck, C.

    A. K., Ed.) Vol. pp89-116, Oxford University Press, Oxford.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A laboratory

    manual, 2nd Ed.*, Cold Spring Harbor Laboratory, Cold Spring Harbor.

Sandhu, G. S., Aleff, R. A. & Kline, B. C. (1992) Dual asymmetric PCR: one-

    step construction of synthetic genes. *BioTech. 12*, 14-16.

Santoro, M. M. & Bolen, D. W. (1988) Unfolding free energy changes

    determined by the linear extrapolation method. 1. Unfolding of

    phenylmethanesulfonyl *a*-chymotrypsin using different denaturants

    Biochemistry 27, 8063-8068.

64

Smith, G. P. & Scott, J. K. (1993) Libraries of peptides and proteins displayed on filamentous phage. *Methods Enzymol. 217*, 228-257.

Smith, G. P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science 228*, 1315-1317.

Smith, C. K., Munson, M. & Regan, L. (1995). Studying α-helix and β-sheet formation in small proteins. Techniques Prot. Chem., 6, 323-332.

Smith, C. K., Withka, J. M. & Regan, L. (1994) A thermodynamic scale for the b-sheet forming tendencies of the amino acids. *Biochemistry 33*, 5510-5517.

Smyth, M. L. & von Itzstein, M. (1994) Design and synthesis of a biologically active antibody mimic based on an antibody-antigen crystal structure. *J. Am. Chem. Soc. 116*, 2725-2733.

Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes *Methods Enzymol. 185*, 60-89.

Suzuki, H. (1994) Recent advances in abzyme studies. *J. Biochem. 115*, 623-628.

Tello, D., Goldbaum, F. A., Mariuzza, R. A., Ysern, X., Schwarz, F. P. & Poljak, R. J. (1993) Immunoglobulin superfamily interactions. *Biochem. Soc. Trans. 21*, 943-946.

Thomas, N. R. (1994) Hapten design for the generation of catalytic antibodies. *Appl. Biochem. Biotech. 47*, 345-372.

Verhoeyen, M., Milstein, C. & Winter, G. (1988) Reshaping human antibodies: Grafting an antilysozyme activity. *Science 239*, 1534-1536.

Venturini et al., (1994) *Protein Peptide Letters 1*, 70-75.

Vuister, G. W. & Bax, A. (1992) Resolution enhancement and spectral editing of uniformly 13C-enriched proteins by homonuclear broadband 13C decoupling. *J. Magn. Reson. 98*, 428-435.

Vuister, G. W., Clore, G. M., Gronenborn, A. M., Powers, R., Garrett, D. S., Tschudin, R. & Bax, A. (1993) Increased resolution and improved spectral quality in four-dimensional 13C/13C-separated HMQC-NOESY-

65

, HMQC spectra using pulsed filed gradients. *J. Magn. Reson. B 101*, 210-213.

Ward, E. S., Güssow, D., Griffiths, A. D., Jones, P. T. & Winter, G. (1989)
Binding activities of a repertoire of single immunoglobulin variable
domains secreted from Escherichia coli *Nature 341*, 554-546.

Webster, D. M., Henry, A. H. & Rees, A. R. (1994) Antibody-antigen
interactions *Curr. Opinion Struct. Biol. 4*, 123-129.

Williams, A. F. & Barclay, A. N., *Ann. Rev. Immunol.* 6:381-405 (1988).

Wilson, I. A. & Stanfield, R. L. (1993) Antibody-antigen interactions. *Curr.
Opinion Struct. Biol. 3*, 113-118.

Wilson, I. A. & Stanfield, R. L. (1994) Antibody-antigen interactions: new
structures and new conformational changes *Curr. Opinion Struct Biol. 4*,
857-867.

Winter, G., Griffiths, A. D., Hawkins, R. E. & Hoogenboom, H. R. (1994)
Making antibodies by phage display technology *Annu. Rev. Immunol. 12*,
433-455.

Wiseman, T., Williston, S., Brandts, J. F. & Lin, L.-N. (1989) Rapid
measurement of binding constants and heats of binding using a new
titration calorimeter. *Anal. Biolchem. 179*, 131-137.

Wittenkind, M. & Mueller, L. (1993) HNCACB, a high-sensitivity 3D NMR
experiment to correlate amide-proton and nitrogen resonances with the
alpha- and beba-carbon resonances in proteins *J. Magn. Reson. B 101*,
201-205.

Wu, T. T., Johnson, G. & Kabat, E. A. (1993) Length distribution of CDRH3 in
antibodies *Proteins: Struct. Funct. Genet. 16*, 1-7.

Wüthrich, K. (1986) NMR of proteins and nucleic acids, John Wiley & Sons,
New York.

Yamazaki, T., Forman-Kay, J. D. & Kay, L. E. (1993) Two-Dimensional NMR
Experiments for Correlating 13C-beta and 1H-delta/epsilon Chemical
Shifts of Aromatic Residues in 13C-Labeled Proteins via Scalar
Couplings. *J. Am. Chem. Soc. 115*, 11054.

Zhang, O., Kay, L. E., Olivier, J. P. & Forman-Kay, J. D. (1994) Backbone 1H and 15N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. *J. Biomol. NMR 4*, 845-858.

67

WHAT IS CLAIMED IS:

1.   A fibronectin type III (Fn3) polypeptide monobody comprising a

plurality of Fn3 β-strand domain sequences that are linked to a plurality

of loop region sequences,

wherein one or more of the monobody loop region sequences vary

by deletion, insertion or replacement of at least two amino acids from the

corresponding loop region sequences in wild-type Fn3, and

wherein the β-strand domains of the monobody have at least a

50% total amino acid sequence homology to the corresponding amino

acid sequence of wild-type Fn3's β-strand domain sequences.

2.   The monobody of claim 1, wherein at least one loop region is capable of

binding to a specific binding partner (SBP) to form a polypeptide:SBP

complex having a dissociation constant of less than 10^{-6} moles/liter.

3.   The monobody of claim 1, wherein at least one loop region is capable of

catalyzing a chemical reaction with a catalyzed rate constant (k_{cat}) and an

uncatalyzed rate constant (k_{uncat}) such that the ratio of k_{cat}/k_{uncat} is greater

than 10.

4.   The monobody of claim 1, wherein one or more of the loop regions

comprise amino acid residues:

i)   from 15 to 16 inclusive in an AB loop;

ii)   from 22 to 30 inclusive in a BC loop;

iii)   from 39 to 45 inclusive in a CD loop;

iv)   from 51 to 55 inclusive in a DE loop;

v)   from 60 to 66 inclusive in an EF loop; and

vi)   from 76 to 87 inclusive in an FG loop.

68

5.   The monobody of claim 1, wherein the monobody loop region sequences vary from the wild-type Fn3 loop region sequences by the deletion or replacement of at least 2 amino acids.

6.   The monobody of claim 1, wherein the monobody loop region sequences vary from the wild-type Fn3 loop region sequences by the insertion of from 3 to 25 amino acids.

7.   An isolated nucleic acid molecule encoding the polypeptide monobody of claim 1.

8.   An expression vector comprising an expression cassette operably linked to the nucleic acid molecule of claim 7.

9.   The expression vector of claim 8, wherein the expression vector is an M13 phage-based plasmid.

10.   A host cell comprising the vector of claim 8.

11.   A method of preparing a fibronectin type III (Fn3) polypeptide monobody comprising the steps of:

a)   providing a DNA sequence encoding a plurality of Fn3 β-strand domain sequences that are linked to a plurality of loop region sequences wherein at least one loop region contains a unique restriction enzyme site;

b)   cleaving the DNA sequence at the unique restriction site;

c)   inserting into the restriction site a DNA segment known to encode a peptide capable of binding to a specific binding partner (SBP) or a transition state analog compound (TSAC) so as to yield a DNA molecule comprising the insertion and the DNA sequence of (a); and

d)   expressing the DNA molecule so as to yield polypeptide
      monobody.

12.   A method of preparing a fibronectin type III (Fn3) polypeptide
      monobody comprising the steps of:

      (a)   providing a replicatable DNA sequence encoding a plurality of
            Fn3 β-strand domain sequences that are linked to a plurality of
            loop region sequences, wherein the nucleotide sequence of at least
            one loop region is known;

      (b)   preparing polymerase chain reaction (PCR) primers sufficiently
            complementary to the known loop sequence so as to be
            hybridizable under PCR conditions, wherein at least one of the
            primers contains a modified nucleic acid sequence to be inserted
            into the DNA;

      (c)   performing polymerase chain reaction using the DNA sequence of
            (a) and the primers of (b);

      (d)   annealing and extending the reaction products of (c) so as to yield
            a DNA product; and

      (e)   expressing the polypeptide monobody encoded by the DNA
            product of (d).

13.   A method of preparing a fibronectin type III (Fn3) polypeptide
      monobody comprising the steps of:

      a)   providing a replicatable DNA sequence encoding a plurality of
            Fn3 β-strand domain sequences that are linked to a plurality of
            loop region sequences, wherein the nucleotide sequence of at least
            one loop region is known;

      b)   performing site-directed mutagenesis of at least one loop region
            so as to create a DNA sequence comprising an insertion mutation;
            and

    c)   expressing the polypeptide monobody encoded by the DNA sequence comprising the insertion mutation.

14.   A kit for performing the method of any one of claims 11-13, comprising a replicatable DNA encoding a plurality of Fn3 β-strand domain sequences that are linked to a plurality of loop region sequences.

15.   A variegated nucleic acid library encoding Fn3 polypeptide monobodies comprising a plurality of nucleic acid species each comprising a plurality of loop regions, wherein the species encode a plurality of Fn3 β-strand domain sequences that are linked to a plurality of loop region sequences,

        wherein one or more of the loop region sequences vary by deletion, insertion or replacement of at least two amino acids from corresponding loop region sequences in wild-type Fn3, and

        wherein the β-strand domain sequences of the monobody have at least a 50% total amino acid sequence homology to the corresponding amino acid sequences of β-strand domain sequences of the wild-type Fn3.

16.   The variegated nucleic acid library of claim 15, wherein one or more of the loop regions encodes:
    i)   an AB amino acid loop from residue 15 to 16 inclusive;
    ii)   a BC amino acid loop from residue 22 to 30 inclusive;
    iii)   a CD amino acid loop from residue 39 to 45 inclusive;
    iv)   a DE amino acid loop from residue 51 to 55 inclusive;
    v)   an EF amino acid loop from residue 60 to 66 inclusive; and
    vi)   an FG amino acid loop from residue 76 to 87 inclusive.

17.   The variegated nucleic acid library of claim 15, wherein the loop region sequences vary from the wild-type Fn3 loop region sequences by the deletion or replacement of at least 2 amino acids.

18. The variegated nucleic acid library of claim 15, wherein the monobody loop region sequences vary from the wild-type Fn3 loop region sequences by the insertion of from 3 to 25 amino acids.

19. The variegated nucleic acid library of claim 15, wherein a variegated nucleic acid sequence comprising from 6 to 75 nucleic acid bases is inserted in any one of the loop regions of said species.

20. The variegated nucleic acid library of claim 15, wherein the variegated sequence is constructed so as to avoid one or more codons selected from the group consisting of those codons encoding cysteine or the stop codon.

21. The variegated nucleic acid library of claim 15, wherein the variegated nucleic acid sequence is located in the BC loop.

22. The variegated nucleic acid library of claim 15, wherein the variegated nucleic acid sequence is located in the DE loop.

23. The variegated nucleic acid library of claim 15, wherein the variegated nucleic acid sequence is located in the FG loop.

24. The variegated nucleic acid library of claim 15, wherein the variegated nucleic acid sequence is located in the AB loop.

25. The variegated nucleic acid library of claim 15, wherein the variegated nucleic acid sequence is located in the CD loop.

26. The variegated nucleic acid library of claim 15, wherein the variegated nucleic acid sequence is located in the EF loop.

72

27.   A peptide display library derived from the variegated nucleic acid library
        of claim 15.

28.   A peptide display library of claim 27, wherein the peptide is displayed on
        the surface of a bacteriophage or virus.

29.   A peptide display library of claim 28, wherein the bacteriophage is M13
        or fd.

30.   A method of identifying the amino acid sequence of a polypeptide
        molecule capable of binding to a specific binding partner (SBP) so as to
        form a polypeptide:SSP complex wherein the dissociation constant of the
        said polypeptide:SBP complex is less than 10^{-6} moles/liter, comprising
        the steps of:

        a)   providing a peptide display library according to claim 28;

        b)   contacting the peptide display library of (a) with an immobilized
                or separable SBP;

        c)   separating the peptide:SBP complexes from the free peptides,

        d)   causing the replication of the separated peptides of (c) so as to
                result in a new peptide display library distinguished from that in
                (a) by having a lowered diversity and by being enriched in
                displayed peptides capable of binding the SBP;

        e)   optionally repeating steps (b), (c), and (d) with the new library of
                (d); and

        f)   determining the nucleic acid sequence of the region encoding the
                displayed peptide of a species from (d) and deducing the peptide
                sequence capable of binding to the SBP.

31.   A method of preparing a variegated nucleic acid library encoding Fn3
        polypeptide monobodies having a plurality of nucleic acid species each
        comprising a plurality of loop regions, wherein the species encode a

plurality of Fn3 β-strand domain sequences that are linked to a plurality

of loop region sequences, wherein one or more of the loop region

sequences vary by deletion, insertion or replacement of at least two

amino acids from corresponding loop region sequences in wild-type Fn3,

and wherein the β-strand domain sequences of the monobody have at

least a 50% total amino acid sequence homology to the corresponding

amino acid sequences of β-strand domain sequences of the wild-type

Fn3, comprising the steps of

a)   preparing an Fn3 polypeptide monobody having a predetermined
        sequence;

b)   contacting the polypeptide with a specific binding partner (SBP)
        so as to form a polypeptide:SSP complex wherein the dissociation
        constant of the said polypeptide:SBP complex is less than 10^{-6}
        moles/liter;

c)   determining the binding structure of the polypeptide:SBP
        complex by nuclear magnetic resonance spectroscopy or X-ray
        crystallography; and

d)   preparing the variegated nucleic acid library, wherein the
        variegation is performed at positions in the nucleic acid sequence
        which, from the information provided in (c), result in one or more
        polypeptides with improved binding to the SBP.

32.   A method of identifying the amino acid sequence of a polypeptide

        molecule capable of catalyzing a chemical reaction with a catalyzed rate

        constant, k_{cat}, and an uncatalyzed rate constant, k_{uncat}, such that the ratio of

        k_{cat}/k_{uncat} is greater than 10, comprising the steps of:

        a)   providing a peptide display library according to claim 28;

        b)   contacting the peptide display library of (a) with an immobilized
                or separable transition state analog compound (TSAC)
                representing the approximate molecular transition state of the
                chemical reaction;

74

c)   separating the peptide:TSAC complexes from the free peptides;

d)   causing the replication of the separated peptides of (c) so as to result in a new peptide display library distinguished from that in (a) by having a lowered diversity and by being enriched in displayed peptides capable of binding the TSAC;

e)   optionally repeating steps (b), (c), and (d) with the new library of (d); and

f)   determining the nucleic acid sequence of the region encoding the displayed peptide of a species from (d) and hence deducing the peptide sequence.

33.   A method of preparing a variegated nucleic acid library encoding Fn3 polypeptide monobodies having a plurality of nucleic acid species each comprising a plurality of loop regions, wherein the species encode a plurality of Fn3 β-strand domain sequences that are linked to a plurality of loop region sequences, wherein one or more of the loop region sequences vary by deletion, insertion or replacement of at least two amino acids from corresponding loop region sequences in wild-type Fn3, and wherein the β-strand domain sequences of the monobody have at least a 50% total amino acid sequence homology to the corresponding amino acid sequences of β-strand domain sequences of the wild-type Fn3, comprising the steps of

a)   preparing an Fn3 polypeptide monobody having a predetermined sequence, wherein the polypeptide is capable of catalyzing a chemical reaction with a catalyzed rate constant, k_{cat}, and an uncatalyzed rate constant, k_{uncat}, such that the ratio of k_{cat}/k_{uncat} is greater than 10;

b)   contacting the polypeptide with an immobilized or separable transition state analog compound (TSAC) representing the approximate molecular transition state of the chemical reaction;

c)   determining the binding structure of the polypeptide:TSAC

complex by nuclear magnetic resonance spectroscopy or X-ray

crystallography; and

d)   preparing the variegated nucleic acid library, wherein the

variegation is performed at positions in the nucleic acid sequence

which, from the information provided in (c), result in one or more

polypeptides with improved binding to or stabilization of the

TSAC.

34.   An isolated polypeptide identified by the method of claim 30.

35.   An isolated polypeptide identified by the method of claim 32.

36.   A kit for identifying the amino acid sequence of a polypeptide molecule

capable of binding to a specific binding partner (SBP) so as to form a

polypeptide:SSP complex wherein the dissociation constant of the said

polypeptide:SBP complex is less than 10^{-6} moles/liter, comprising the

peptide display library of claim 28.

37.   A kit for identifying the amino acid sequence of a polypeptide molecule

capable of catalyzing a chemical reaction with a catalyzed rate constant,

k_{cat}, and an uncatalyzed rate constant, k_{uncat}, such that the ratio of k_{cat}/k_{uncat}

is greater than 10, comprising the peptide display library of claim 28.

38.   A polypepetide derived by using the kit of claim 36.

39.   A polypeptide derived by using the kit of claim 37.

FIG. 1A
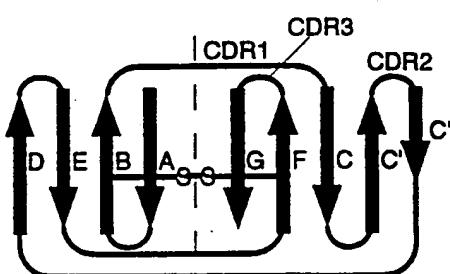
Immunoglobulin VH

FIG. 1B
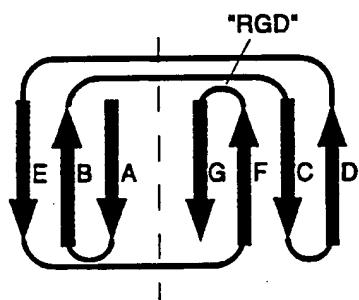
Fibronectin type III

FIG. 1C

FIG. 1D

```
NdeI                          PstI                  EcoRI
 1        11        21        31        41
mq VSDVPRDLEV VAATPTSLLI SWDAPAVTVR YYRITYGETG GNSPVQEFTV
          A         B                  C              D

              SalI              SacI        XhoI
51        61        71        81        91
  PGSKSTATIS GLKPGVDYTI TVYAVTGRGD SPASSKPISI NYRT
          E                 F                 G
```

FIG. 2

FIG. 3A

FIG. 3B

FIG. 4A

FIG. 4B
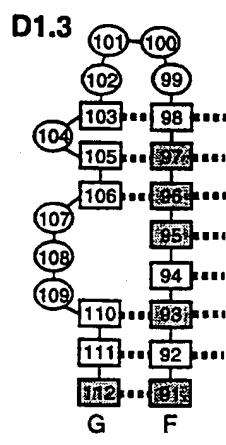
FIG. 4C
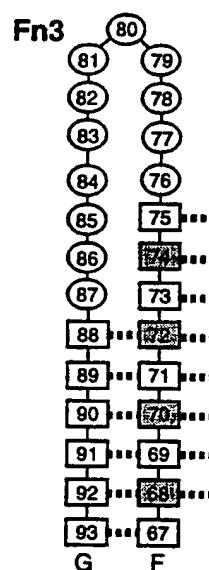
FIG. 4D

NdeI
CATATGCAGGTTTCTGATGTTCCGCGTGACCTGGAAGTTGTTGCTGCGACCCCGACTAGC
    MetGlnValSerAspValProArgAspLeuGluValValAlaAlaThrProThrSer
     -2 -1 1   10

    BclI PvuII   PstI   BsiWI
CTGCTGATCAGCTGGGATGCTCCTGCAGTTACCGTGCGTTATTACCGTATCACGTACGGT
LeuLeuIleSerTrpAspAlaProAlaValThrValArgTyrTyrArgIleThrTyrGly
         20   30

                            EcoRI
GAAACCGGTGGTAACTCCCCGGTTCAGGAATTCACTGTACCTGGTTCCAAGTCTACTGCT
GluThrGlyGlyAsnSerProValGlnGluPheThrValProGlySerLysSerThrAla
         40   50

                        SalI   Bst1107I
ACCATCAGCGGCCTGAAACCGGGTGTCGACTATACCATCACTGTATACGCTGTTACTGGC
ThrIleSerGlyLeuLysProGlyValAspTyrThrIleThrValTyrAlaValThrGly
         60   70

                    SacI   XhoI
CGTGGTGACAGCCCAGCGAGCTCCAAGCCAATCTCGATTAACTACCGTACCTAGTAACTC
ArgGlyAspSerProAlaSerSerLysProIleSerIleAsnTyrArgThr
         80   90

FIG. 5

FIG. 6

FIG. 7

FIG. 8

phage ELISA (ubiquitin)

FIG. 9

competition phage ELISA

FIG. 10

FIG. 11

Phagemid ELISA of
Fluorescein Binders

FIG. 12

Phagemid ELISA
Fluorescein Competition
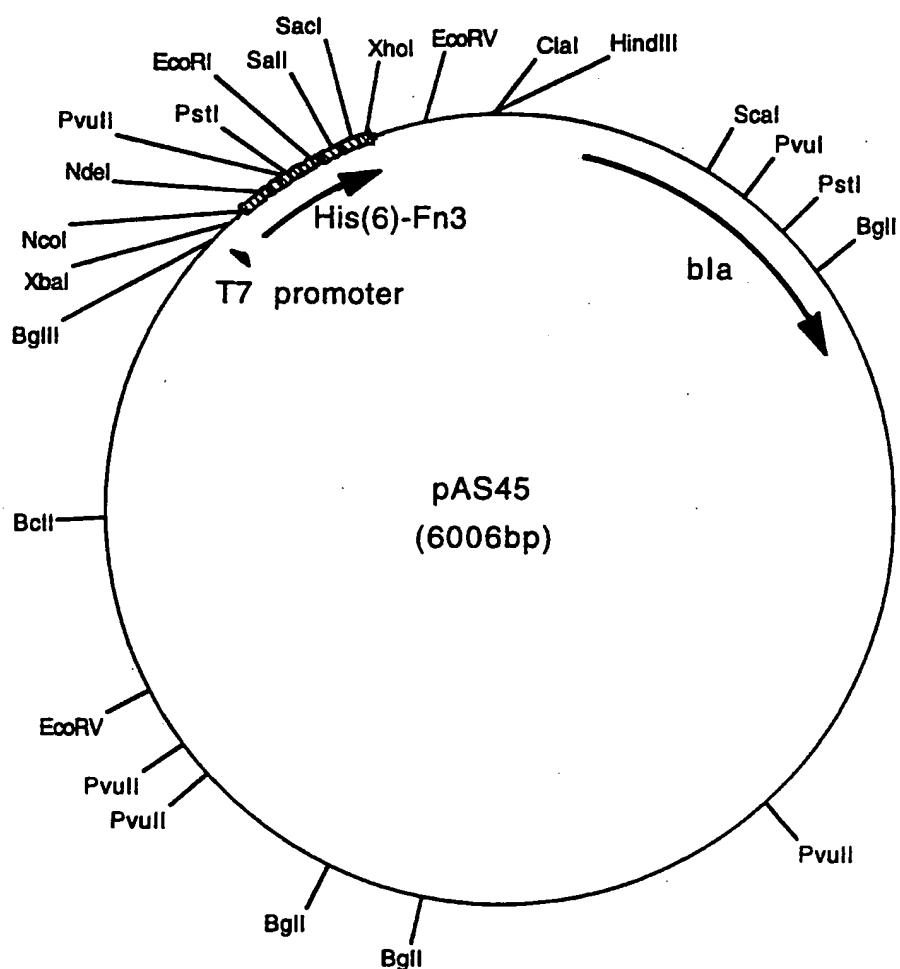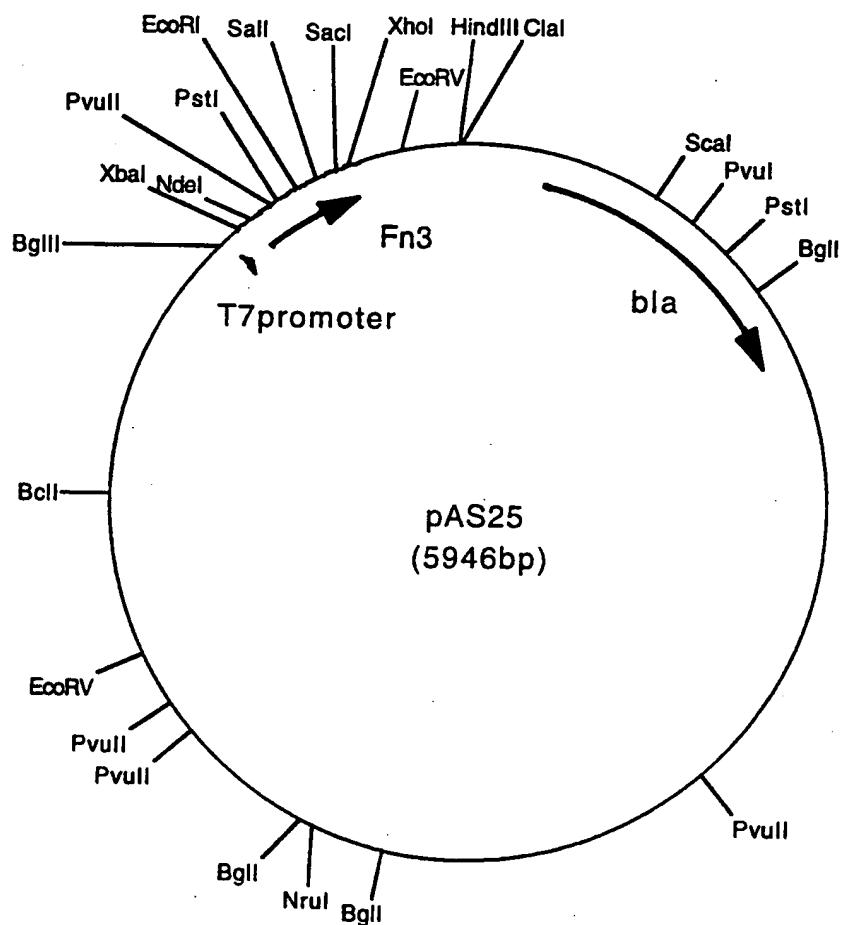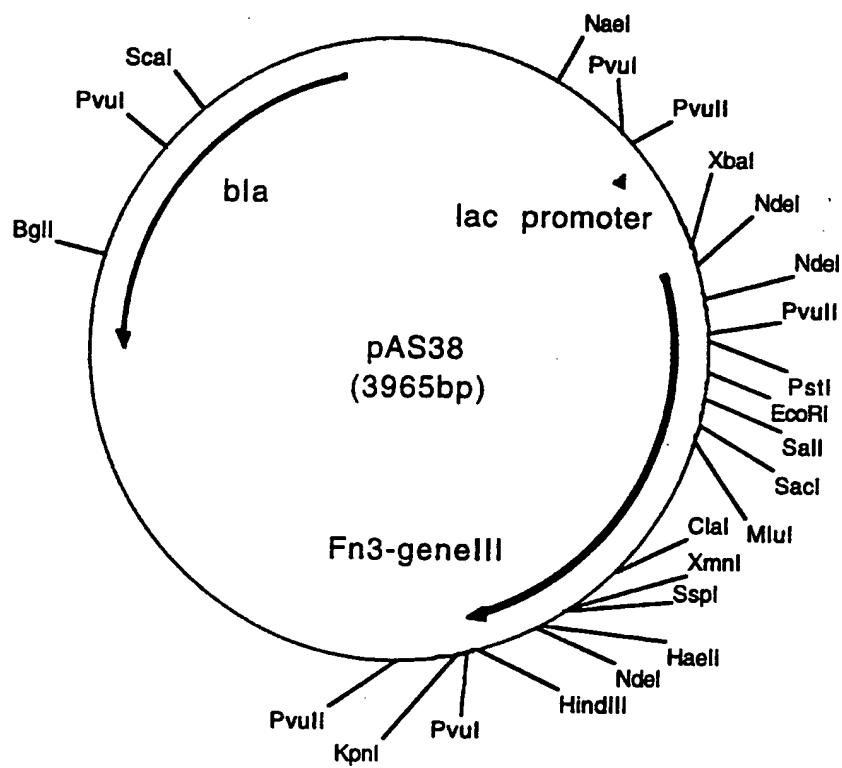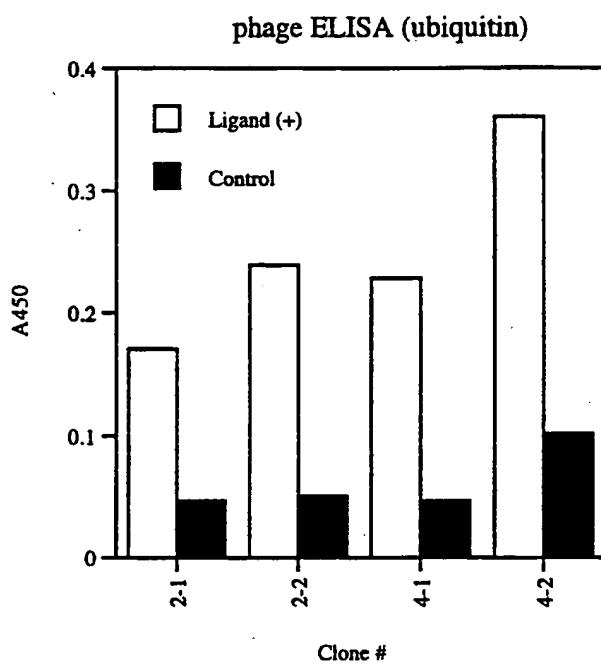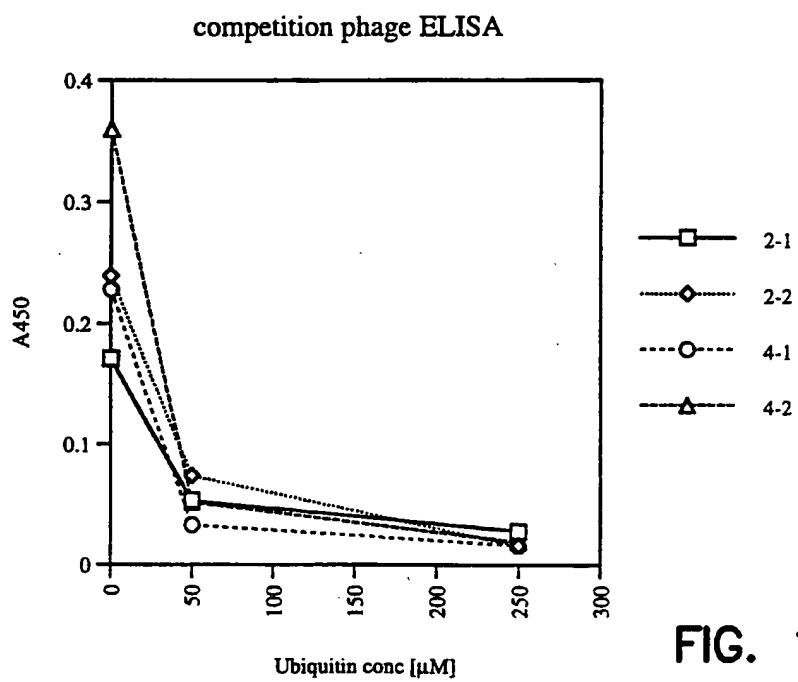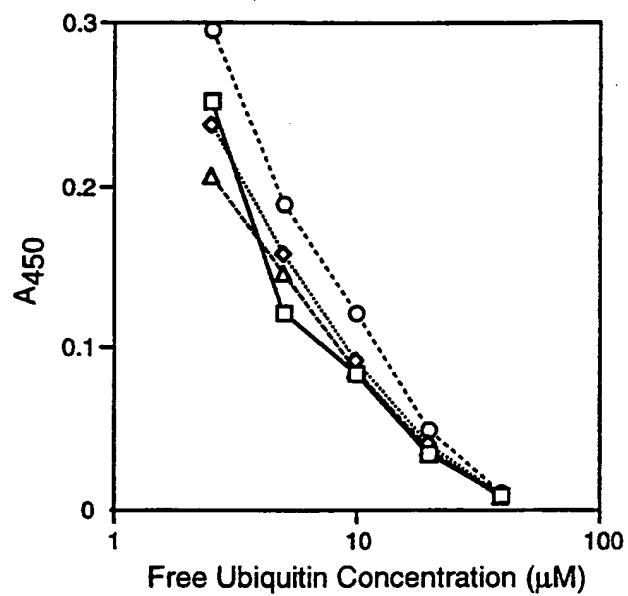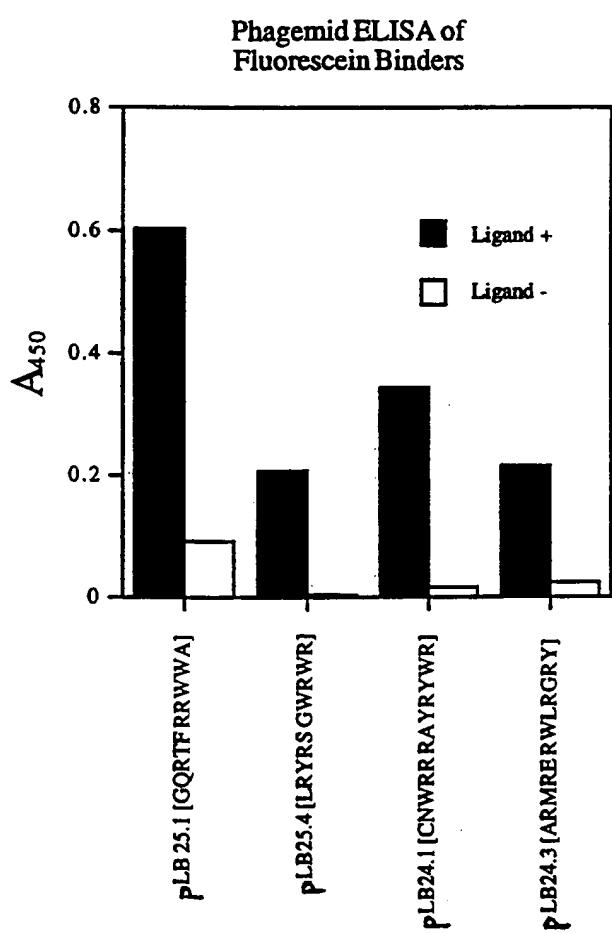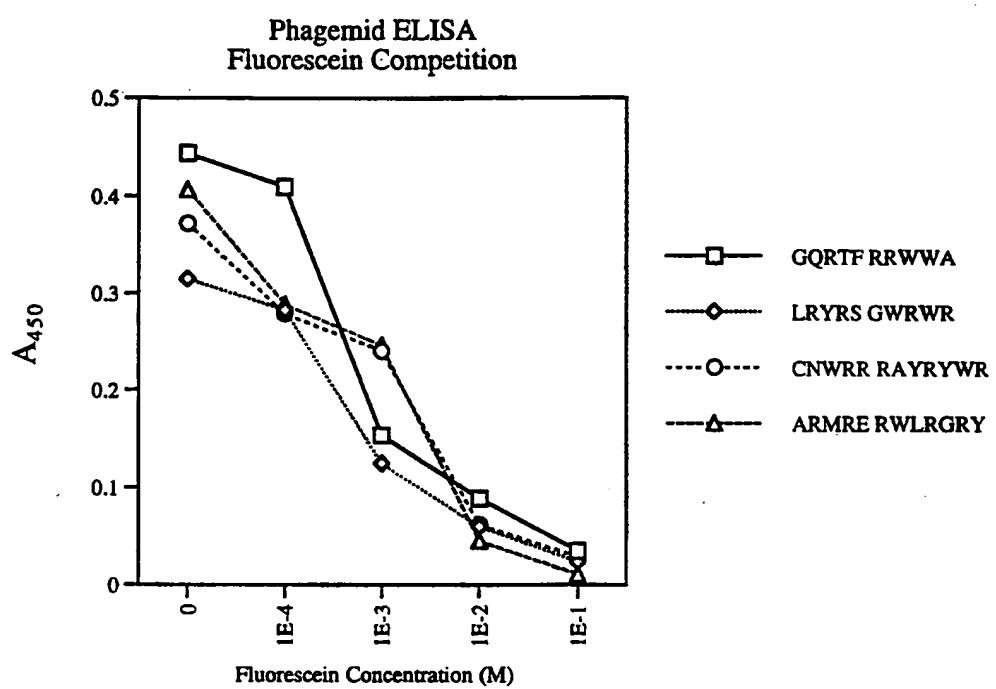
FIG. 13
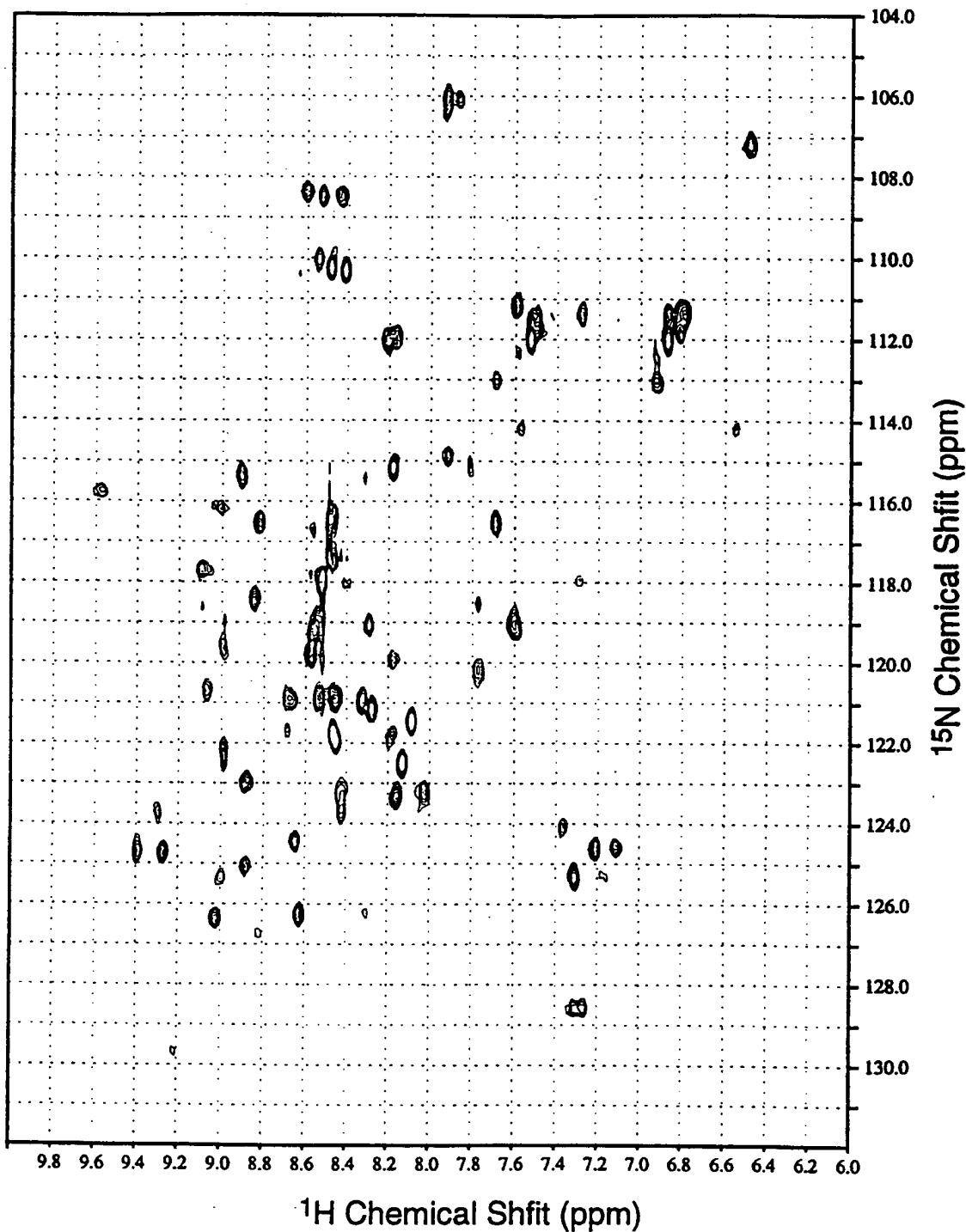
FIG. 14
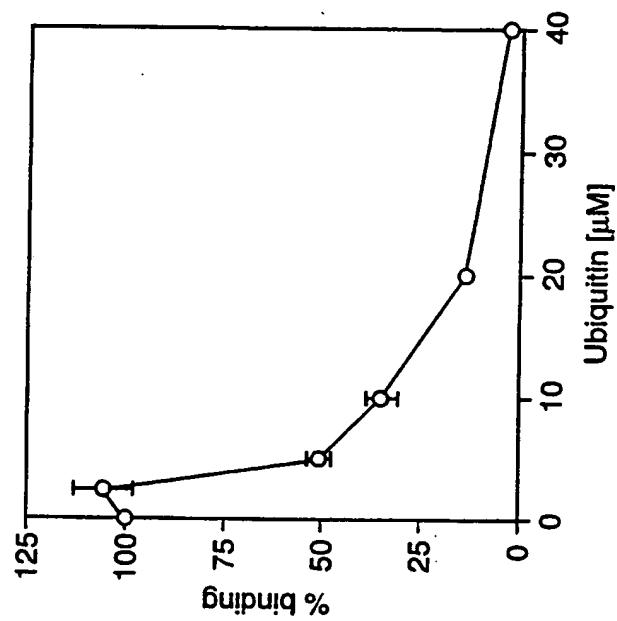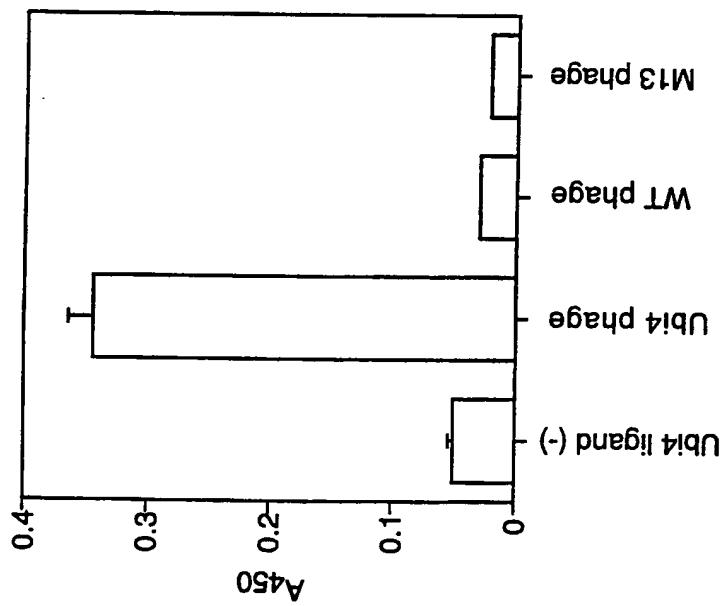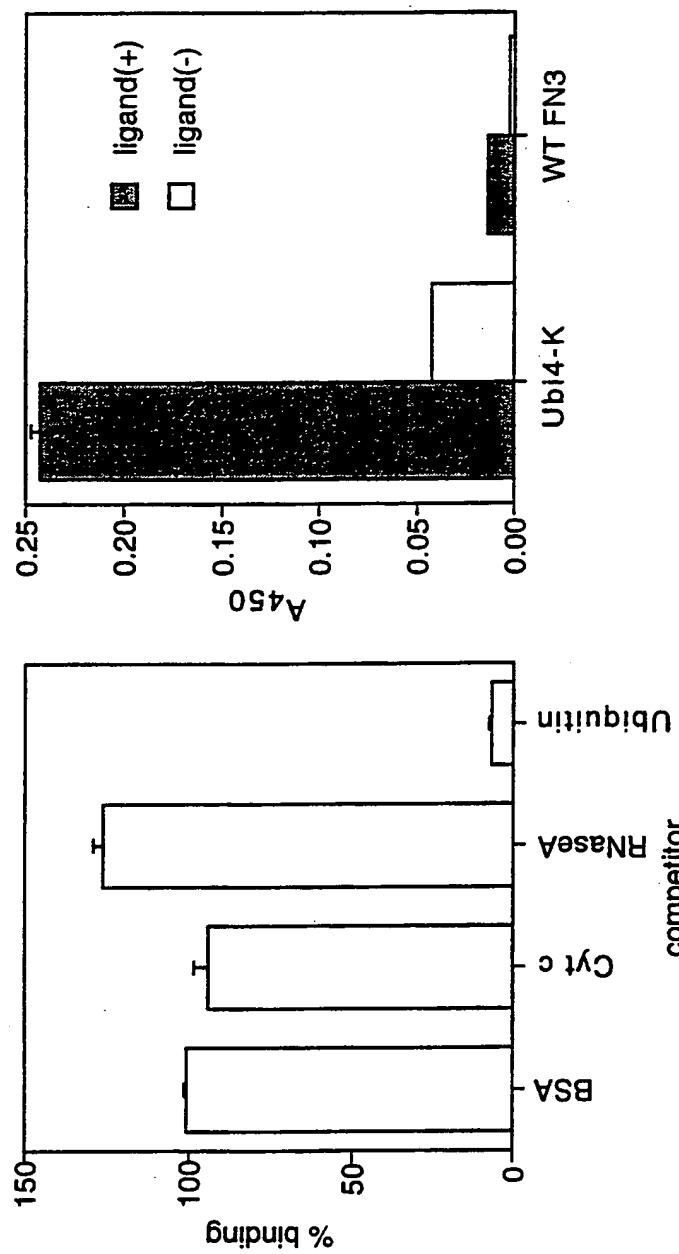
FIG. 15B

FIG. 15A

FIG. 15D

FIG. 15C
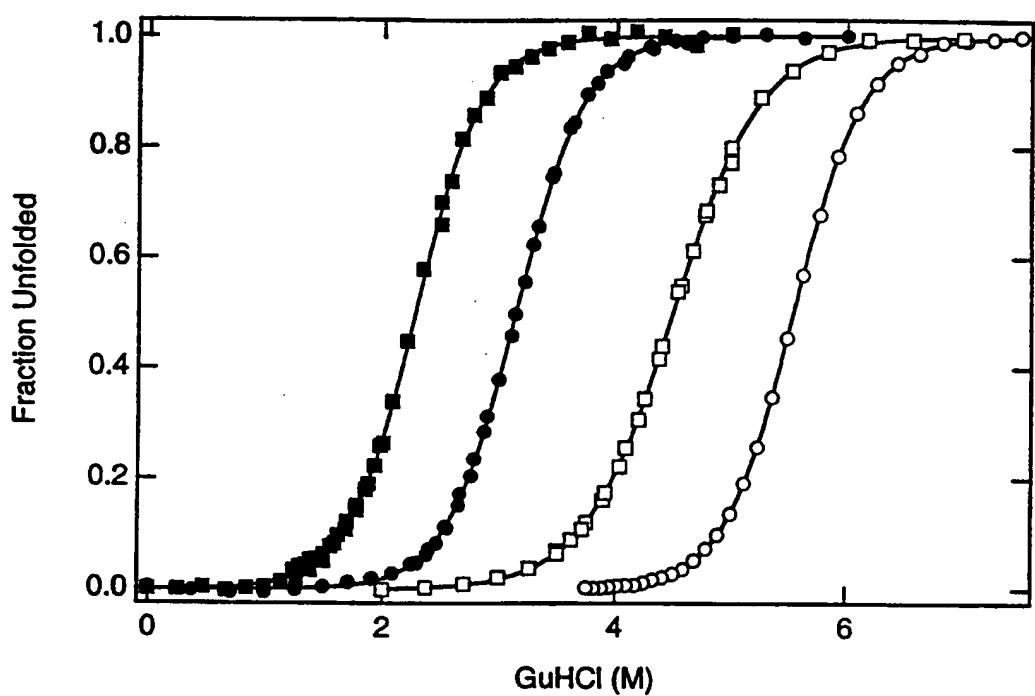
FIG. 16
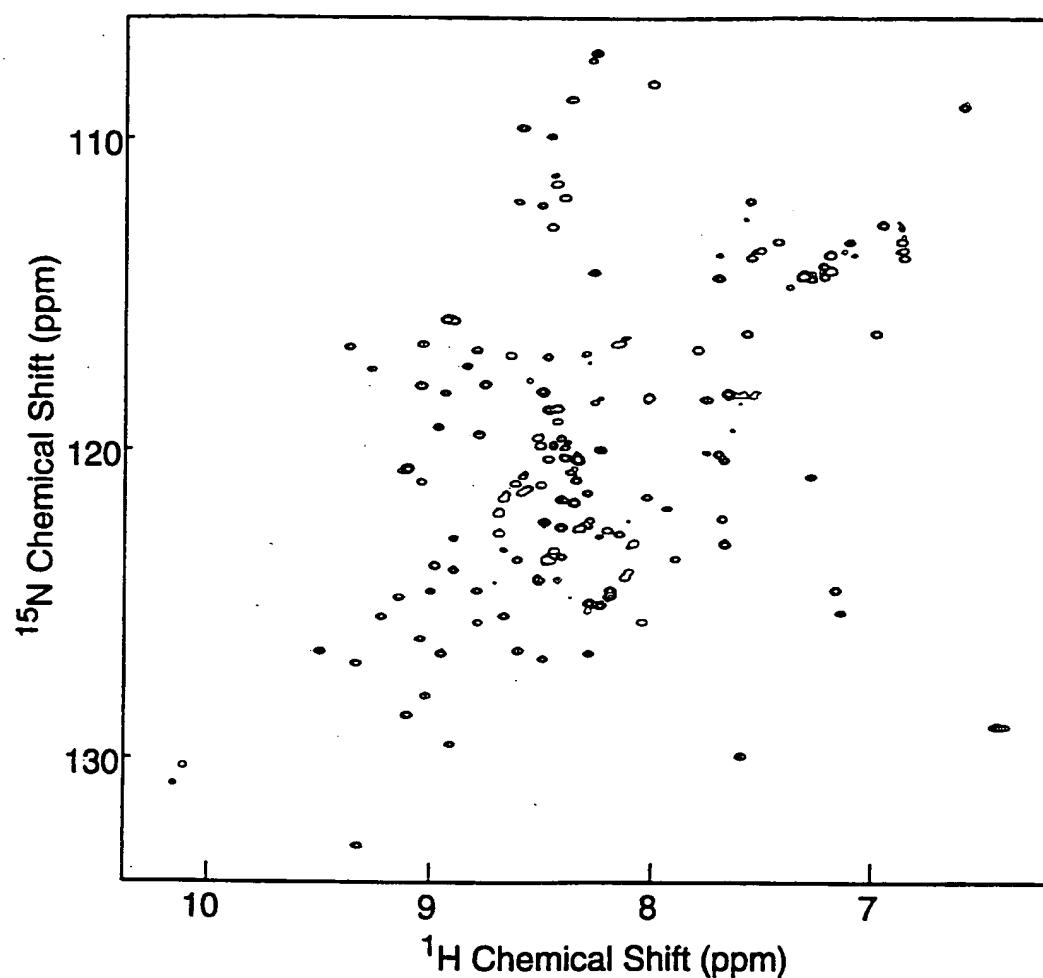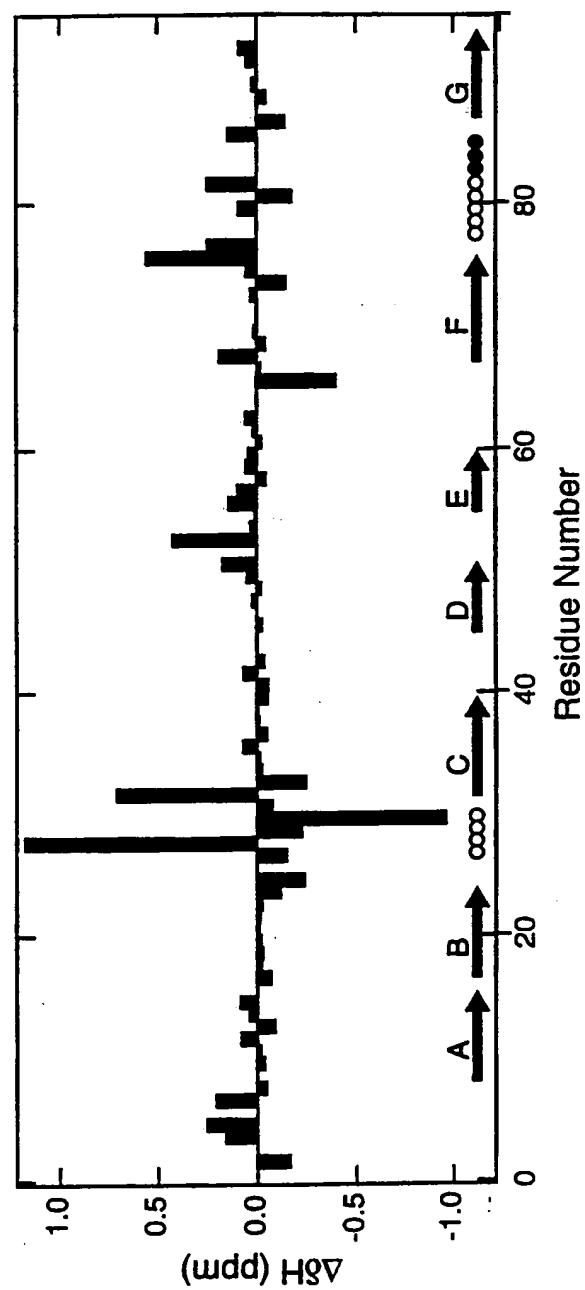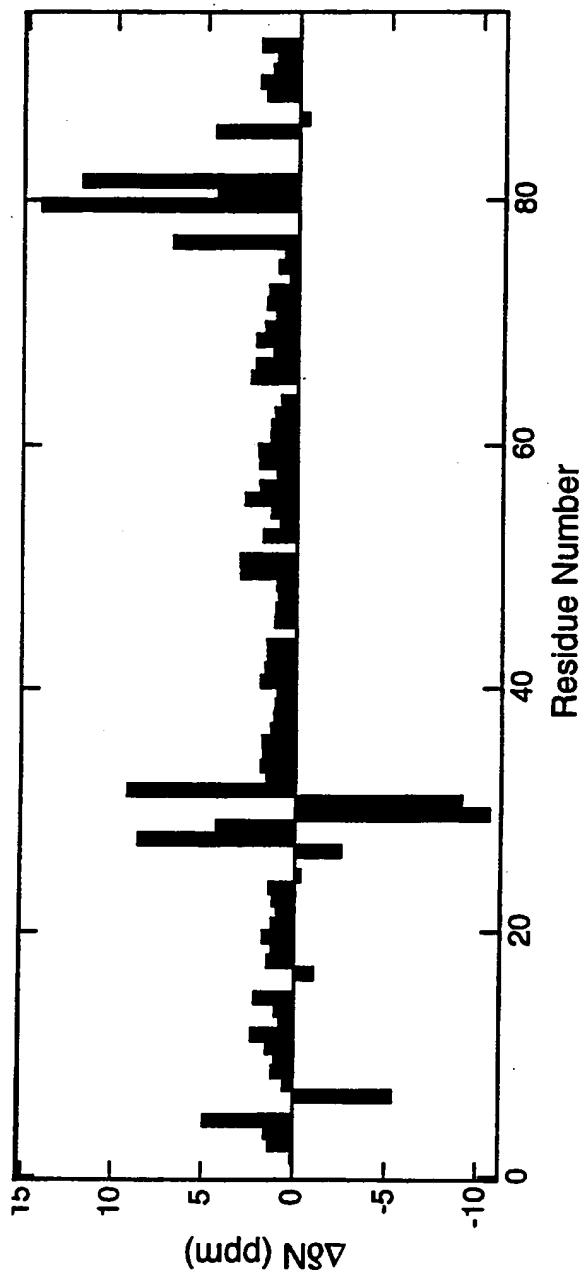
FIG. 17A

FIG. 17B

FIG. 17C

Rational Modification of Protein Stability by the Mutation of Charged Surface Residues[†]

Shari Spector,[‡,§] Minghui Wang,[||] Stefan A. Carp,[⊥] James Robblee,[○] Zachary S. Hendsch,[⊥] Robert Fairman,[○]
Bruce Tidor,[*,⊥] and Daniel P. Raleigh[*,||,⊗]

*Department of Physiology and Biophysics, State University of New York at Stony Brook, Stony Brook, New York 11794-8661,
Department of Chemistry, State University of New York at Stony Brook, Stony Brook, New York 11794-3400, Department of
Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, Department of Molecular, Cellular
and Developmental Biology, Haverford College, Haverford Pennsylvania 19041, and Graduate Programs in Biophysics and in
Molecular and Cellular Biology, State University of New York at Stony Brook, Stony Brook, New York 11794*

Received September 7, 1999; Revised Manuscript Received November 3, 1999

ABSTRACT: Continuum methods were used to calculate the electrostatic contributions of charged and polar side chains to the overall stability of a small 41-residue helical protein, the peripheral subunit-binding domain. The results of these calculations suggest several residues that are destabilizing, relative to hydrophobic isosteres. One position was chosen to test the results of these calculations. Arg8 is located on the surface of the protein in a region of positive electrostatic potential. The calculations suggest that Arg8 makes a significant, unfavorable electrostatic contribution to the overall stability. The experiments described in this paper represent the first direct experimental test of the theoretical methods, taking advantage of solid-phase peptide synthesis to incorporate approximately isosteric amino acid substitutions. Arg8 was replaced with norleucine (Nle), an amino acid that is hydrophobic and approximately isosteric, or with α-amino adipic acid (Aad), which is also approximately isosteric but oppositely charged. In this manner, it is possible to isolate electrostatic interactions from the effects of hydrophobic and van der Waals interactions. Both Arg8Nle and Arg8Aad are more thermostable than the wild-type sequence, testifying to the validity of the calculations. These replacements led to stability increases at 52.6 °C, the T_m of the wild-type, of 0.86 and 1.08 kcal mol^{-1}, respectively. The stability of Arg8Nle is particularly interesting as a rare case in which replacement of a surface charge with a hydrophobic residue leads to an increase in the stability of the protein.

The amino acid sequences of proteins include a wide variety of different residue types, including several acidic and basic groups. The charges on the surface of a protein are certainly important for its solubility, but what effect do electrostatic interactions have on the overall stability of the molecule? A number of recent experimental and theoretical studies have suggested that partially or completely buried salt bridges function at least in part to provide specificity to

the fold, although they do not generally provide added stability beyond that of a hydrophobic bridge of similar geometry (*1, 2*). This conclusion is based on results showing that the favorable electrostatic interactions from the salt bridge are often insufficient to overcome the electrostatic desolvation penalty (*3−8*). Surface salt bridges appear to make only small contributions to protein stability (*9−12*). However, in T4 lysozyme, a partially exposed salt bridge appears to contribute 3−5 kcal mol^{-1} to the stability of the protein (*13*). Experimental studies have also been performed to examine the contribution of a single charged residue to the stability of a protein. The binding face of barstar, the inhibitor of the ribonuclease barnase, has four acidic residues. Replacement of any of these with alanine leads to an increase in the stability of the protein. On the basis of the ionic strength dependence, the increased stability of the barstar mutants is ascribed to the removal of unfavorable electrostatic interactions (*14*).

Comparison of these results is complicated by the choice of different reference states. In the T4 lysozyme study, the salt bridge in the wild-type protein is only partially exposed, and the mutation cycle involves changing each member of the salt bridge pair to asparagine, together and individually (*13*). In one barnase study investigating an existing, solvent-exposed salt bridge triad, a triple mutant cycle is used in which each residue is substituted with alanine (*10*). In another

[*] To whom correspondence should be addressed. Bruce Tidor: telephone, 617-253-7258; fax, 617-252-1816; e-mail, tidor@mit.edu. Daniel Raleigh: telephone, 516-632-9547; fax, 516-632-7960; e-mail, draleigh@notes.cc.sunysb.edu.
[‡] Department of Physiology and Biophysics, State University of New York at Stony Brook.
[§] Current address: Departments of Chemistry and Biology, Massachusetts Institute of Technology 68-565, 77 Massachusetts Avenue, Cambridge, MA 02139-4307.
[||] Department of Chemistry, State University of New York at Stony Brook.
[⊥] Department of Chemistry, Massachusetts Institute of Technology.
[○] Department of Molecular, Cellular and Developmental Biology, Haverford College.
[⊗] Graduate Programs in Biophysics and in Molecular and Cellular Biology, State University of New York at Stony Brook.

barnase study, the wild-type Asp12/Thr16 pair is compared to Ala/Thr, Asp/Arg, and Ala/Arg (9). Finally, in a third study of barnase, a salt bridge is engineered into an existing helix. In this case, the salt-bridging pair is compared to the wild-type sequence in a double mutant cycle in which Ser28 is replaced by glutatmate and Ala32 is replaced by lysine (11). Such mutant cycles assume that the interaction energy between the salt-bridging residues is purely electrostatic. It is possible for there to be hydrophobic and van der Waals interactions between residues in the wild-type sequence or in any of the mutants in the cycle, and different reference states make it difficult to compare the results of different studies. Moreover, conformational changes between mutants complicate the analysis further. These additional interactions are not accounted for in studies in which only single mutations are made.

Calculations using continuum electrostatics can estimate the contribution of a single charged residue to the overall stability of a protein. For example, such calculations can compare the stability of the protein with its native sequence to a variant in which an individual charged residue has been replaced by a hydrophobic isostere (15, 16). In this manner, all other interactions within the protein are kept constant, and only the electrostatic interactions formed by a single residue with each of the other charged and polar groups in the protein are considered. Even so, there are a large number of interactions to account for in the calculations. For such calculations to be tractable, it is helpful to choose smaller proteins as models.

This study focuses on the peripheral subunit-binding domain, derived from the dihydrolipoamide acetyltransferase component (EC 2.3.1.12) of the pyruvate dehydrogenase multienzyme complex from *Bacillus stearothermophilus*. Because of its small size, 43 amino acids, it is an attractive target for such calculations. It adopts a stable, unique tertiary fold in the absence of any disulfide bridges or ligand binding (17–19). Its structure is comprised of two parallel alpha helices connected by a loop containing a short stretch of 3_{10}-helix (Figure 1). The loop is maintained in a unique conformation via a hydrogen bonding network between a buried, charged aspartate residue, Asp34, near the N-terminus of the second helix and several of the backbone amides in the loop. The variant used in the calculations and experiments described here is 41 amino acids long, corresponding to residues 3–43 of the peripheral subunit-binding domain (18, 19) and will be referred to as the wild-type protein or psbd41.[1]

The small size of the peripheral subunit-binding domain offers another distinct advantage. The calculations compare the stability of variants in which the acidic and basic residues are replaced by hydrophobic isoteres. This is not often experimentally testable using natural amino acids, but some

---
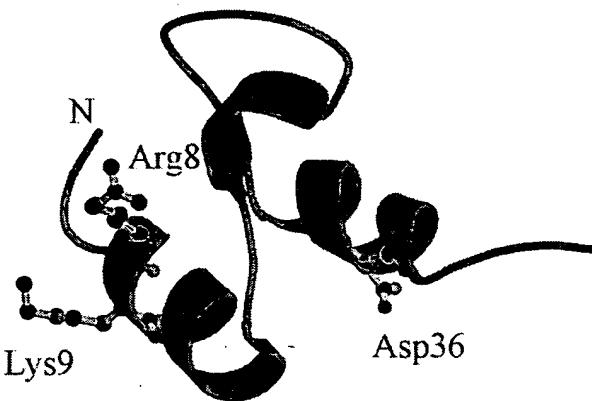
1   11   21
VIAMPSVRKY AREKGVDIRL VQGTGKNGRV
31   41
LKEDIDAFLA GGA

FIGURE 1: Molscript diagram (44) and sequence of the peripheral subunit-binding domain. The side chains of Arg8, Lys9, and Asp36 are displayed on the ribbon diagram, the N-terminus of the protein is labeled, and the position of Arg8 in the sequence is emphasized in bold.

unnatural amino acids are close approximations. Since these proteins cannot be produced in high yield using traditional expression systems, solid-phase peptide synthesis must be used to prepare the large quantities of such variants required for biophysical characterization. The relatively short sequence of psbd41 means that it can easily be prepared by solid-phase peptide synthesis. The use of hydrophobic isosteres is expected to maintain all nonelectrostatic interactions between the residue of interest and the remainder of the protein and allows the isolation of electrostatic contributions to the stability of the protein. The use of isosteric amino acid substitutions makes this study the first direct experimental test of the theoretical methods.

In this paper, we report the results of a set of continuum electrostatic calculations performed on the peripheral subunit-binding domain. The results of the calculations suggest that there are several residues located on the surface of the protein that provide a significant unfavorable electrostatic contribution to the overall stability of the domain, in part due to the asymmetry of electrostatic potential mapped to the surface of the protein. In particular, we have chosen Arg8 as a test case for the calculation. Replacement of Arg8 with norleucine, which approximates a hydrophobic isostere, leads to a significant increase in the thermal stability of the peripheral subunit-binding domain. Substitution of Arg8 with α-amino adipic acid, which is roughly isosteric but oppositely charged, leads to a further increase in thermal stability. The results of this study suggest a general strategy for increasing the stability of a protein by minimizing unfavorable surface interactions.

MATERIALS AND METHODS

Materials. Fmoc-PAL-PEG-PS resin was purchased from Perseptive Biosystems (Foster City, CA). HOBt and HBTU were purchased from Advanced ChemTech (Louisville, KY). Fmoc-L-α-amino adipic acid-δ-*tert*-butyl ester was from

Bachem Bioscience Inc. (King of Prussia, PA). All other Fmoc-protected amino acids were purchased from Perseptive Biosystems and Advanced ChemTech. D_2O and (trimethylsilyl)-proprionate were obtained from Cambridge Isotope Laboratories, Inc. (Andover, MA). All other solvents and reagents were obtained from Fisher Scientific (Springfield, NJ).

Calculations. The best representative (20) from the family of NMR structures of the peripheral subunit-binding domain, Protein Data Bank identifier 2pdd (17), was used for the calculations. Hydrogen atoms were placed using the HBUILD algorithm (21) in CHARMM (22) with standard pH 7 titration states for all amino acid side chains. Continuum electrostatic calculations were carried out with a modified version of the DELPHI computer program (23−25) using our previously published methods (26), except where differences are noted below. The PARSE parameter set (27) was used with protein and solvent dielectric constants of 4 and 80, respectively, a temperature of 300 K, and solvent ionic strength of 68 mM, corresponding to the experimental conditions. The protein−solvent boundary was defined as the analytic molecular surface of the protein with no dielectric smoothing applied. Computations were carried out for charging each amino acid side chain individually to permit the estimation of the electrostatic desolvation and the interaction contributions for each side chain.

Peptide Synthesis and Purification. Peptides were prepared by solid-phase synthesis using a Millipore 9050 Plus automated peptide synthesizer and standard Fmoc chemistry. Arg8Nle and Arg8Aad correspond to residues 3−43 of the peripheral subunit-binding domain in which Arg8 has been replaced by norleucine or α-amino adipic acid. Both peptides are N-terminally acetylated and C-terminally amidated. The peptides were purified by HPLC on a C18 reverse phase column (Vydac) in two steps. The solvent system used in the first step was a water−acetonitrile gradient containing 170 mM triethylamine phosphate. The second step used a water−acetonitrile gradient containing 0.1% (v/v) trifluoroacetic acid. Both peptides were greater than 95% pure as judged by HPLC.

The identity of each peptide was confirmed by matrix-assisted laser desorption and ionization time-of-flight mass spectrometry (MALDI-TOF). Arg8Nle had an experimental weight of 4383.1 Da (expected 4386.0), and Arg8Aad had an experimental weight of 4420.7 Da (expected 4416.2).

Analytical Utracentrifugation. Analytical ultracentrifugation was performed to test whether Arg8Nle and Arg8Aad are monomeric. Each sample was dialyzed against 2 mM phosphate, 2 mM borate, 2 mM citrate, 50 mM NaCl. Equilibrium experiments were performed at 25 °C with a Beckman Optima XL-A analytical ultracentrifuge using rotor speeds of 30 000, 40 000, and 50 000 rpm. Six-channel, 12 mm path length, charcoal-filled Epon cells with quartz windows were used. Ten scans were averaged. Partial specific volumes were calculated from the weighted average of the partial specific volumes of the individual amino acids and solution densities were calculated using standard tables listing coefficients for the power series approximation of density (28). This calculation was compared to a gravimetric determination of the solution system used. The HID program from the Analytical Ultracentrifugation Facility at the University of Connecticut was used for data analysis.

Circular Dichroism. All circular dichroism (CD) experiments described here were carried out on an Aviv 62A DS circular dichroism spectrophotometer using a buffer containing 2 mM sodium phosphate, 2 mM sodium borate, 2 mM sodium citrate, and 50 mM sodium chloride at pH 8.0. The protein concentrations for all experiments were obtained by measuring the absorbance at 276 nm in 6 M GdnHCl, 20 mM NaH_2PO_4, pH 6.5, using an extinction coefficient of 1450 M^{-1} cm^{-1}. The concentration dependence of the CD signal at 25 °C was monitored for both Arg8Nle and Arg8Aad at 222 nm, and the mean residue ellipticity was shown to be independent of concentration for all of our experimental conditions.

NMR Spectroscopy. All NMR experiments were performed on either a Varian Instruments Inova 500 MHz or Inova 600 MHz nuclear magnetic resonance spectrometer. The peptides were dissolved in 90% H_2O, 10% D_2O, pH 5.4, with (trimethylsilyl)- proprionate as a chemical shift standard. The concentrations of Arg8Nle and Arg8Aad were 1 mM and 4 mM, respectively. One-dimensional NMR spectra were acquired at 25 °C using standard presaturation methods. Two-dimensional data sets were also collected at 25 °C. For both Arg8Nle and Arg8Aad, TOCSY (total correlation spectroscopy)(29, 30) and NOESY (two-dimensional nuclear Overhauser enhancement spectroscopy)(31, 32) spectra were acquired with a spectral width of 6000.6 Hz on the 500 MHz spectrometer. The mixing times were 75 ms for the TOCSY and 250 ms for the NOESY. The collected data sets, with matrix sizes of 512 × 2048, were processed with Felix95.0 (Molecular Simulations Inc., 1995) on an SGI Indigo[2] workstation. All chemical shift assignments were made using standard procedures (33).

Thermal Denaturations. Thermal denaturations were monitored by far-UV CD (222 nm) and near-UV CD (280 nm) in a stirred 1 cm cuvette. The temperature was raised in 2 degree intervals from 2 to 98 °C for Arg8Nle and from 2 to 90 °C for Arg8Aad. The sample was allowed to equilibrate for 1.2 min, and the signal was averaged for 45 s. Reversibility was confirmed by comparing the ellipticity at 2 °C after a thermal denaturation to the initial ellipticity at 2 °C. Thermal denaturations of Arg8Nle were greater than 97% reversible, and for Arg8Aad, they were greater than 99% reversible. All thermal denaturations were analyzed by nonlinear least squares curve fitting using SigmaPlot (Jandel Scientific) as described previously (18, 19). Data are normalized to fraction unfolded. The errors in the thermodynamic parameters were analyzed using an F-test to determine the 95% confidence limits (19, 34).

RESULTS

Calculations. The analysis of the continuum electrostatic calculations is presented in Table 1. The electrostatic desolvation and intraprotein interaction contributions to the free energy of folding are listed for each polar or charged side chain. All of the side chains are computed to have essentially zero or net unfavorable electrostatic effects on the folding of the peripheral subunit-binding domain. The only exception is Asp17, which is buried in the structure and makes hydrogen bonds with the backbone NH groups of Arg19 and Leu20, as well as additional favorable electrostatic interactions with other neighboring groups. The

Table 1: Electrostatic Free Energies in the Peripheral Subunit-binding Domain[a]

sidechain	desolvation penalty (kcal mol^{-1})	interaction free energy (kcal mol^{-1})	total free energy (kcal mol^{-1})
Met4	0.1	0.4	0.5
Pro5	0.0	0.2	0.2
Ser6	1.8	1.0	2.8
Arg8	1.9	1.9	3.8
Lys9	1.9	2.5	4.4
Tyr10	1.5	0.3	1.8
Arg12	0.3	1.1	1.4
Glu13	0.9	−1.0	−0.1
Lys14	0.5	−1.0	−0.5
Asp17	3.3	−6.5	−3.2
Arg19	2.0	0.2	2.1
Gln22	0.3	0.4	0.6
Thr24	1.9	−0.7	1.2
Lys26	1.4	−1.0	0.4
Asn27	2.1	−0.7	1.4
Arg29	0.3	0.3	0.6
Lys32	0.2	−0.7	−0.5
Glu33	1.1	0.5	1.6
Asp34	11.2	−10.7	0.5
Asp36	1.5	0.6	2.1
Phe38	0.6	0.2	0.8

[a] The desolvation penalty describes the difference in solvation free energy of a residue in the native versus denatured state. The interaction free energy describes the energetics of the electrostatic interactions between a residue and the remainder of the protein. The total free energy is the sum of the desolvation penalty and the interaction free energy.

total effect of Asp17 is computed to be favorable by 3.2 kcal mol^{-1}. The most surprising result of the calculations, however, is the large number of amino acid side chains computed to have *unfavorable* interactions in the folded structure of the protein. One generally expects *favorable* folded state interactions that are offset by unfavorable desolvation penalties. Examination of the structure reveals that these repulsions include a grouping of positively charged residues near the surface, Arg8, Lys9, and Arg12, which lie along the exposed face of an α-helix. Figure 2 shows the resulting electrostatic potential. We are cautious about interpreting the results of calculations based on the best representative from a family of NMR structures determined for a small protein whose structure, particularly at the protein surface, is likely to be fluctuating. For this reasons, we feel that averaging over many conformations, which is beyond the scope of the current report, may improve the accuracy of the values in Table 1. Nevertheless, it is reasonable to expect unfavorable effects for the positive surface cluster because each side chain is clearly partially desolvated and there are certainly repulsions among the members of the set.

Amino Acid Substitutions. The calculations suggest that three residues, Arg8, Lys9, and Asp36, make a significant unfavorable electrostatic contribution to the overall stability of the peripheral subunit-binding domain. Of these, Arg8 was chosen to test the results of the calculations. It is located on the surface of the peripheral subunit-binding domain in a region of strong positive electrostatic potential (Figure 2). Arg8 is the second residue in a helix, and therefore, its charged guanidino group may interact unfavorably with backbone dipolar groups in the helix. In addition, there is another arginine on the same face of the helix at position 12, four residues away from Arg8, and these two residues could also interact unfavorably. Replacement of Arg8 with a hydrophobic residue should eliminate these unfavorable

electrostatic interactions. Substitution with a negatively charged residue could provide further stability through favorable salt-bridge and backbone dipole interactions, as well as through other interactions with the local positive potential.

On this basis, Arg8 was replaced both with a hydrophobic residue and with a negatively charged residue, each similar to arginine in size and shape. For the hydrophobic substitution, Arg8 was replaced with norleucine (Arg8Nle), which has a straight chain aliphatic side chain four carbons long, an arginine analogue with a methyl group in place of the guanidino group. For the substitution of opposite charge, α-amino adipic acid (Arg8Aad) was chosen. This unnatural amino acid also has the same number of methylene groups as arginine, but the terminal guanidino group is replaced by a carboxylate.

Surface Charge Variants of Psbd41 are Monomeric. Both Arg8Nle and Arg8Aad remain monomeric throughout the concentration range of the experiments reported here. It was especially important to test this for Arg8Nle, since replacement of a surface charge with a hydrophobic residue could result in a sticky patch prone to association or aggregation. However, Arg8Nle is monomeric. The molar ellipticity at 222 nm is independent of concentration over the range 24−460 μM. Furthermore, the one-dimensional NMR spectrum is also identical for samples at 440 μM and 3.8 mM. If aggregation were to occur, broadening of the NMR lines would be expected, but the lines remain sharp at the higher concentration. Analytical ultracentrifugation also shows that Arg8Nle is monomeric. For a 242 μM sample, a single species fit gave a molecular weight of 4600 ± 200 Da, compared to a calculated molecular weight of 4386.0 Da, consistent with Arg8Nle remaining monmeric at this concentration. Fits with multiple species models were no better than the single species fit, as judged by the randomness of the residuals (data not shown). Finally, thermal denaturations were performed at 12 μM and 582 μM (see below), resulting, within the experimental uncertainty, in identical thermal denaturation midpoints and identical values of the enthalpy at the midpoint of the transition. All of these taken together provide strong evidence that Arg8Nle remains monomeric over the concentration range of interest.

Arg8Aad is also monomeric. The molar ellipticity at 222 nm is independent of concentration over the studied range of 58 μM to 1.42 mM. The one-dimensional NMR spectra of 500 μM and 4 mM Arg8Aad are also identical, which would be unlikely if aggregation were occurring. Analytical ultracentrifugation results confirm that at 343 μM Arg8Aad is monomeric. Using a single species analysis results in a molecular weight of 4600 ± 200 Da, compared to the calculated molecular weight of 4416.2 Da, and there is no improvement in the fit using multiple species models (data not shown). Finally, the midpoint of the thermal denaturation and the enthalpy at the midpoint are identical for 15 μM and 440 μM samples of Arg8Aad, providing additional evidence that the protein remains monomeric over the concentration range used for the experiments in this paper.

Arg8Nle and Arg8Aad adopt the same structure as wild-type psbd41. Spectroscopic evidence suggests that both Arg8Nle and Arg8Aad adopt essentially the same fold as the wild-type protein. The near-UV CD spectra have the same shape, and the signal at 280 nm at 25 °C is similar for the
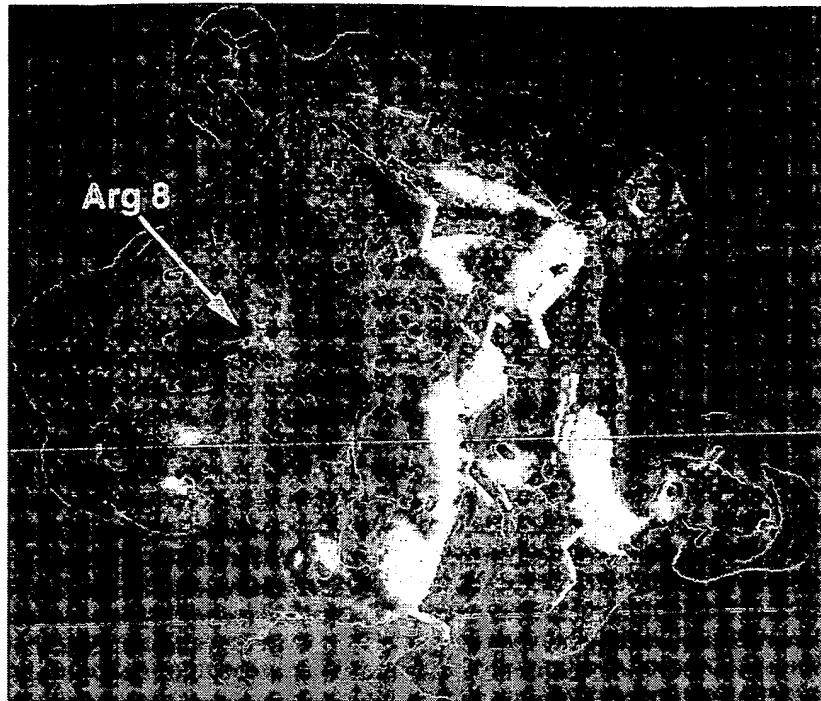
FIGURE 2: GRASP figure of the peripheral subunit-binding domain, shown in the same orientation as in Figure 1. Negative and positive values of electrostatic potential are indicated by linearly deepening shades of red and blue, respectively.

three proteins, with values of 75.9 deg cm^2 $dmol^{-1}$ for Arg8Nle, 104.1 deg cm^2 $dmol^{-1}$ for Arg8Aad, and 119.5 deg cm^2 $dmol^{-1}$ for the wild-type. The far-UV CD spectra are also almost identical in shape, with $[\Theta]_{222}$ values at 25 °C of $-10\,600$ deg cm^2 $dmol^{-1}$ for Arg8Nle, $-11\,000$ deg cm^2 $dmol^{-1}$ for Arg8Aad, and $-11\,100$ deg cm^2 $dmol^{-1}$ for wild-type (data not shown).

Stronger evidence that the three proteins adopt the same structure comes from their one- and two-dimensional NMR spectra and their chemical shift assignments. The one-dimensional NMR spectra all show several very characteristic peaks. First are the two sets of ring-current-shifted methyl protons from Val16 and Val21, which are both clearly present in the three spectra (Figure 3A). For the wild-type protein, Val16 and Val21 appear at 0.42 and 0.23 ppm, respectively. They resonate at 0.41 and 0.23 ppm in Arg8Nle, and 0.42 and 0.23 ppm in Arg8Aad. Another characteristic resonance that is well-resolved in the one-dimensional spectrum arises from the amide proton of Thr24, which is hydrogen bonded to the buried, charged aspartate residue at position 34. As a result, it appears significantly downfield at 9.98 ppm in the wild-type protein. In Arg8Nle, this proton resonates at 9.94 ppm, and in Arg8Aad, its chemical shift is also 9.94 ppm. If the structure of Arg8Nle or Arg8Aad were significantly different from that of the wild-type protein, much larger differences in these chemical shifts would be expected.

Nearly complete resonance assignments were possible for both surface charge variants and are available from the authors upon request. The NMR assignments for the peptide backbone provide additional evidence that the surface charge variants adopt the same structure as the wild-type protein. For Arg8Nle, none of the assigned $C^\alpha H$ resonances has a chemical shift that differs from the wild-type protein by more than 0.1 ppm. In contrast, the $C^\alpha H$ chemical shifts differ from random coil values (*35*) by between -0.62 and 0.42

ppm (Figure 3B). Similarly, only one amide chemical shift, that of Met4, differs from the wild-type assignments by more than 0.20 ppm, and this residue is near the N-terminus. The assignments for Arg8Aad also agree well with those for the wild-type protein. Only two of the assigned C^α protons have a chemical shift different from the wild-type by more than 0.06 ppm (Figure 3C). In contrast, the C^α proton chemical shifts in Arg8Aad differ from average random coil values by between -0.41 and $+0.64$ ppm. The NH chemical shifts are also quite similar for Arg8Aad and wild-type psbd41, with only two residues having chemical shifts that differ by more than 0.11 ppm. Since the chemical shift of a proton is so sensitive to its environment, it is highly unlikely that the proteins could have such similar NMR spectra if they adopted different structures.

Stability Measurements. Thermal denaturation of the peripheral subunit-binding domain and the two surface charge variants shows that substitution of Arg8 results in an increase in thermal stability for both variants (Figure 4). In addition, all three proteins undergo two-state folding, as evidenced by the excellent agreement between the values of T_m and $\Delta H°$ (T_m) for thermal denaturations monitored by near- and far-UV CD spectroscopy. Averaged values are reported in Table 2. The uncertainties given here are determined by F-value analysis of the nonlinear regressions. Since such errors are often asymmetric, we report the larger limit on the error. The wild-type protein has a T_m of 53.1 ± 0.8 °C by far-UV CD and 52.1 ± 1.4 °C by near-UV CD, while $\Delta H°$ (T_m) is 33.4 ± 3.3 kcal mol^{-1} by far-UV CD and 30.1 ± 5.0 kcal mol^{-1} by near-UV CD. Replacement of Arg8 with norleucine results in an increase in the T_m to 61.9 ± 1.1 °C by far-UV CD and 61.4 ± 1.1 °C by near-UV CD. The $\Delta H°$ (T_m) is essentially unchanged, with values of 33.9 ± 4.4 kcal mol^{-1} determined by far-UV CD, and 33.7 ± 4.2 kcal mol^{-1} determined by near-UV CD. Arg8Aad is
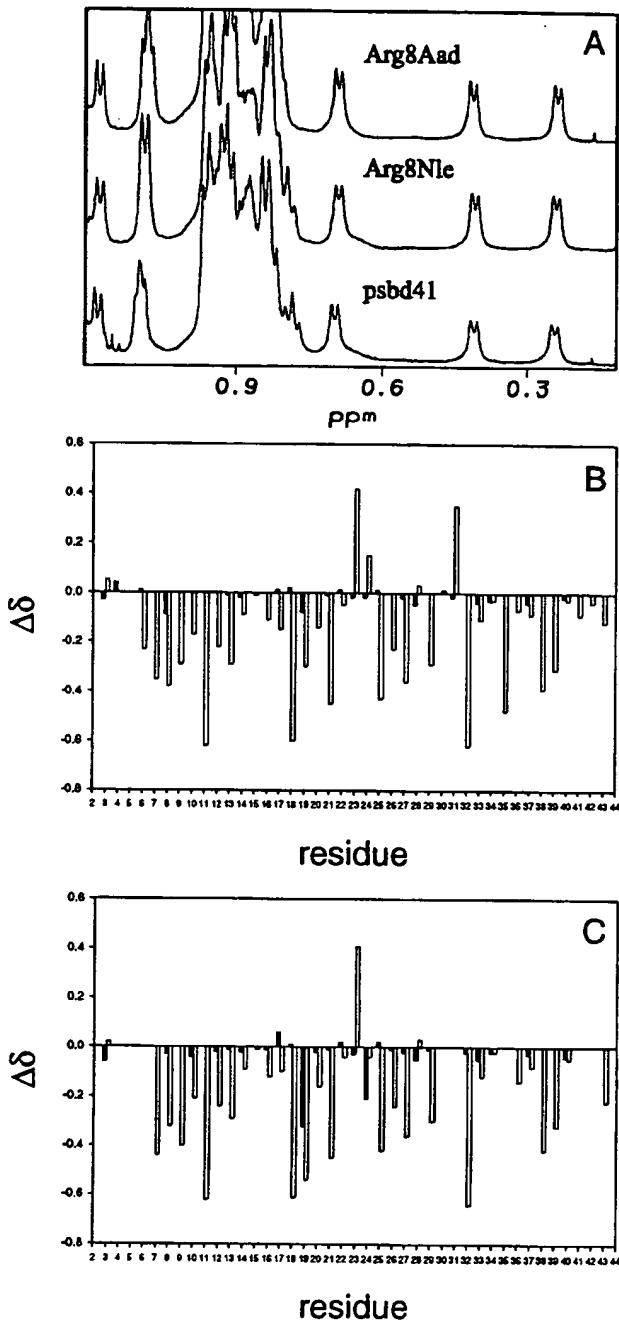
FIGURE 3: (A): One-dimensional NMR spectra of psbd41 (bottom), Arg8Nle (center), and Arg8Aad (top). (B, C): C$^\alpha$H chemical shift differences between the surface charge variants and psbd41 (filled bars) or random coil values (open bars) (35). B, Arg8Nle; C, Arg8Aad.

further stabilized, relative to Arg8Nle, with a T_m of 64.7 ± 1.7 °C by far-UV CD and 64.2 ± 2.1 °C by near-UV CD. The $\Delta H°$ (T_m) remains similar to the wild-type and to Arg8Nle, with values of 34.9 ± 4.6 kcal mol^{-1} determined by far-UV CD, and 31.5 ± 4.3 kcal mol^{-1} determined by near-UV CD.

Thermal denaturation experiments are analyzed using the Gibbs—Helmholtz equation. To do this requires knowledge of the heat capacity change, $\Delta C°_p$. We used the value determined previously for the peripheral subunit-binding domain, 0.43 kcal mol^{-1} K^{-1} (19), not only for the analysis of the data from the wild-type protein, but also for the two
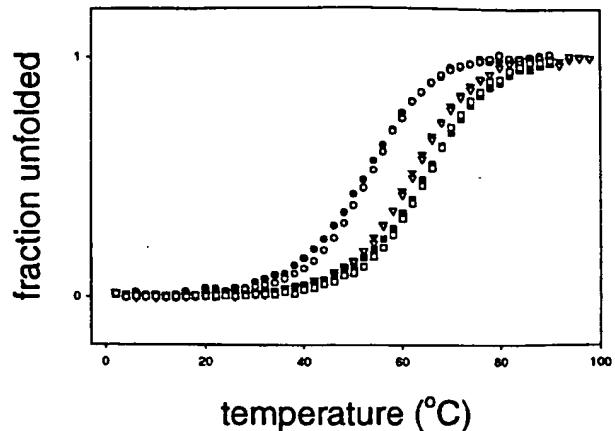

FIGURE 4: Thermal denaturations. Open symbols correspond to far-UV CD data collected at 222 nm, and solid symbols correspond to near-UV CD data collected at 280 nm. O: 14.5 μM psbd41. ●: 435 μM psbd41. ▽: 12 μM Arg8Nle. ▼: 582 μM Arg8Nle. □: 15 μM Arg8Aad. ■: 440 μM Arg8Aad.

surface charge variants, and we make the additional assumption that $\Delta C°_p$ is independent of temperature. The heat capacity change is related to the difference in accessible surface area between the native and denatured states (36). On the basis of the NMR and CD data, Arg8Nle and Arg8Aad both adopt the same tertiary structure as the wild-type protein. Thus, they should have the same heat capacity change as the wild-type, unless the mutations perturb the structure of the denatured state ensemble. For Arg8Aad, this is unlikely, since the carboxylate on α-amino adipic acid will remain charged and solvent exposed at pH 8. However, the norleucine in Arg8Nle could potentially participate in a non-native hydrophobic cluster in the unfolded state, which might affect the heat capacity. The expected effect on $\Delta C°_p$ would be small, probably at least an order of magnitude less than the uncertainty in the measured value (0.43 ± 0.25 kcal mol^{-1} K^{-1}). The magnitude of the uncertainty results from the value being at the lower limit of what is measurable. The heat capacity could not be determined by differential scanning microcalorimetry, because the transition was too broad for an accurate analysis and instead was determined by the global analysis of thermal denaturations performed in the presence of a variety of chemical denaturant concentrations (19). The difference between the measured value and that for the wild-type is certain to be less than the uncertainty in the numbers. In addition, analysis of the thermal denaturations of the wild-type protein using a variety of values of $\Delta C°_p$ does not significantly affect the results. The T_m of psbd41 is insensitive to the value of the heat capacity used in the analysis, and the enthalpy at the midpoint does not change, within the uncertainty in the measurement, using heat capacities ranging from 0.30 to 0.65 kcal mol^{-1} K^{-1} (18, 19). The value of $\Delta C°_p$ determined for the wild-type, therefore, provides a reasonable estimate of the heat capacity of the surface charge variants.

Using the value of $\Delta C°_p$ for the wild-type protein and the measured values of $\Delta H°$ (T_m) and T_m from the surface charge variants, we have calculated $\Delta\Delta G°_{D-N}$ at the T_m of the wild-type protein. At 52.6 °C, $\Delta\Delta G°_{D-N}$ is equal to 0.86 kcal mol^{-1} for Arg8Nle, and 1.08 kcal mol^{-1} for Arg8Aad (Table 2). Evaluating the equation at 27 °C, the temperature assumed in the calculations, leads to $\Delta\Delta G°_{D-N}$ values of 0.65 kcal

Table 2: Thermodynamic Properties of the Peripheral Subunit-binding Domain and the Surface Charge Variants

	$T_m{}^a$ (°C)	ΔH° (T_m) (kcal mol^{-1})	$\Delta G^\circ{}_{D-N}$(27 °C)^b (kcal mol^{-1})	$\Delta\Delta G^\circ{}_{D-N}$(27 °C) (kcal mol^{-1})	$\Delta\Delta G^\circ{}_{D-N}$(52.6 °C) (kcal mol^{-1})
psbd41	52.6	31.8	2.06		
Arg8Nle	61.7	33.8	2.71	0.65	0.86
Arg8Aad	64.5	33.2	2.76	0.70	1.08

^{a} T_m and ΔH°(T_m) are the average of the values obtained from the analysis of thermal denaturations monitored by near- and far-UV CD. ^{b} $\Delta G^\circ{}_{D-N}$(27 °C) and $\Delta G^\circ{}_{D-N}$(52.6 °C) are obtained using the Gibbs–Helmholtz equation, the values of T_m and ΔH°(T_m) from this table.

mol^{-1} for Arg8Nle and 0.70 kcal mol^{-1} for Arg8Aad (Table 2). However, estimating the stability at 27 °C requires a long extrapolation and leads to greater uncertainty in the values of $\Delta\Delta G^\circ{}_{D-N}$. Calculations using the range of ΔH° (T_m) and $\Delta C^\circ{}_p$ suggested by the uncertainties in the parameters result in only small changes in the stability at 53 °C, because the extrapolation is, at most, 11°, whereas there is much greater variation at 27 °C. Thus, as predicted by the calculations, substitution of Arg8 with a hydrophobic residue of approximately the same size and shape results in an increase in the stability of the protein. The difference in stability between the surface charge variants and the wild-type protein is greater at the T_m of the wild-type than at 27 °C, and these values are also included in Table 2.

Under ideal circumstances, the stabilities of the three proteins should also be compared by measurement using chemical denaturation. However, for this set of proteins there was no good choice of chemical denaturant. The two most common denaturants are guanidine hydrochloride and urea. It is not possible to obtain accurate thermodynamic parameters from a urea denaturation of the peripheral subunit-binding domain, because it is so small. Like the heat capacity, the m-value of chemical denaturation also depends on the difference in accessible surface area between the native and denatured states (*36*). Because psbd41 is so small, it has a low m-value, and therefore, very broad transitions. Guanidine hydrochloride, on the other hand, is a salt. Since ionic strength affects the stability of psbd41 (*19*), and because we are interested in electrostatic effects, guanidine is also not a suitable denaturant.

DISCUSSION

The continuum electrostatic methods used here work well to predict the role an individual charged residue plays in the stability of a protein. Arg8 is predicted to be destabilizing by 3.8 kcal mol^{-1}. Substitution of Arg8 with a hydrophobic residue of similar size and shape leads to an increase in stability at 27 °C of 0.65 kcal mol^{-1}, which is significant but much smaller than predicted. Part of this discrepancy could be due to the somewhat smaller size of norleucine relative to arginine. The difference in surface area buried could be responsible for roughly 1.5 kcal mol^{-1}, or half the total discrepancy (*37*). Static-structure continuum calculations with low internal dielectric may somewhat overestimate the size of mutational effects on stability (*38*). While increasing the value used for the internal dielectric may improve the agreement with experiment in some instances, it is more likely that explicitly sampling conformational degrees of freedom will be more appropriate. At 52.6 °C, the T_m of the wild-type, $\Delta\Delta G^\circ{}_{D-N}$ is 0.86 kcal mol^{-1}. The replacement of a surface charge with a hydrophobic residue would not normally be expected to increase the stability of the protein,

because it is more favorable for a hydrophobic side chain to be buried within the core of the protein, rather than exposed to solvent. However, because of its unusual environment the charge on Arg8 is unfavorable, and in this case the substitution leads to an increase in stability. This is in contrast to observations made with λ Cro protein, in which substitution of a tyrosine on the surface of the protein with smaller hydrophobic residues or with polar or charged residues leads to an increase in its stability. This phenomenon was dubbed the reverse hydrophobic effect because the residue in question becomes more exposed in the native structure than it is in the unfolded state (*39*). Obviously, any reverse hydrophobic effect occurring in Arg8Nle is compensated by the removal of unfavorable electrostatic interactions. The calculations do not take into account any changes in the denatured state, and a reverse hydrophobic effect would be expected to stabilize the denatured state of the hydrophobic mutant. This may account for some of the discrepancy between the calculated and experimental $\Delta\Delta G^\circ{}_{D-N}$ values. In addition, the calculations assume a single static structure and do not take into account side chain dynamics. This may also contribute to the difference between the observed and calculated stability changes.

Substitution of Arg8 with α-amino adipic acid leads to a further increase in stability at 27 °C, with a $\Delta\Delta G^\circ{}_{D-N}$ of 0.70 kcal mol^{-1} relative to wild-type, or 1.08 kcal mol^{-1} at 52.6 °C. Calculations similar to those described here suggest Arg8Aad to be about 1.5 kcal/mol more stable than Arg8Nle (data not shown); again, the correct direction but an overestimate of the magnitude. While in Arg8Nle a number of unfavorable electrostatic interactions are alleviated, Arg8Aad can overcome the unfavorable exposure of the norleucine side chain to solvent and may make several favorable electrostatic interactions with the remainder of the protein. The α-amino adipic acid side chain has the potential to interact favorably with helix backbone dipolar groups, and to form a salt bridge with Arg12.

Another interesting residue described in the calculations is Asp34. This residue is more than 95% buried, and its side chain takes part in hydrogen bonds to the backbone amides of Gly23, Thr24, Gly25, and Leu31 in the loop, and to the side chain hydroxyl group of Thr24. A buried charge is normally expected to be extremely unfavorable due to the large desolvation penalty (*3*). The calculations suggest that in this case the burial of Asp34 is only modestly destabilizing by 0.6 kcal mol^{-1}. This is likely due to the extensive hydrogen bonding network. Our previously reported studies of Asn and Val substitutions at this position have demonstrated that Asp34 is important for the specificity of the fold and also contributes to the stability of the domain (*18, 40*). These studies are in qualitative agreement with the calculations.

The amino acid changes chosen for this study have a distinct advantage, arising from the use of hydrophobic isosteres. It is difficult to choose a reference state such that only electrostatic interactions are considered. Although norleucine and α-amino adipic acid are not perfectly isosteric with arginine, they serve as excellent approximations. Through these changes we were able to overcome unfavorable electrostatic interactions and increase the stability of the peripheral subunit-binding domain.

The substitutions were made in the context of a model study. Nevertheless, it is instructive to consider their effect in a biological context. The peripheral subunit-binding domain is a piece of a much larger enzyme, the dihydrolipoamide acetyltransferase (E2). The domain's function within the pyruvate dehydrogenase multienzyme complex is to direct intermolecular interactions with each of the other two enzymes in the complex, the pyruvate decarboxylase (E1, EC 1.2.4.1) and the dihydrolipoamide dehydrogenase (E3, EC 1.8.1.4). The molecular details of the interactions between the peripheral subunit-binding domain and E1 have not been characterized, but there is a crystal structure available of the peripheral subunit-binding domain bound to E3 (*41*). In the complex, Arg8 is important for binding, forming a salt bridge with Glu431 of E3. Although the substitutions described in this paper would not serve this protein well in vivo, the methodology could nevertheless be applied to other proteins if care is taken to avoid residues involved in catalysis or intermolecular interactions.

Relatively little attention has been paid to the contributions of surface electrostatic interactions to the stability of globular proteins. This study provides a clear demonstration that alleviating unfavorable surface interactions can increase the stability of proteins. Many proteins contain clusters of positively or negatively charged residues, and the results presented here suggest that optimization of surface electrostatic interactions is likely to be a generally applicable strategy for enhancing protein stability. For example, in recent studies of ribonuclease T1 and ubiquitin, it was shown that relieving surface charge repulsion through mutation increased protein stability (*42, 43*). These methods may prove useful, for example, in structural studies of marginally stable proteins, since surface mutations are much less likely to perturb the structure than a mutation to the core, or in membrane-associated proteins.

REFERENCES

1. Lumb, K. J., and Kim, P. S. (1995) *Biochemistry 34*, 8642–8648.
2. Raleigh, D. P., Betz, S. F., and DeGrado, W. F. (1995) *J. Am. Chem. Soc. 117*, 7558–7559.
3. Hendsch, Z. S., and Tidor, B. (1994) *Protein Sci. 3*, 211–226.
4. Honig, B., and Yang, A.-S. (1995) *Adv. Prot. Chem. 46*, 27–58.
5. Sindelar, C. V., Hendsch, Z. S., and Tidor, B. (1998) *Protein Sci. 7*, 1898–1914.
6. Waldburger, C. D., Schildbach, J. F., and Sauer, R. T. (1995) *Nat. Struct. Biol. 2*, 122–128.
7. Wang, L., O'Connell, T., Tropsha, A., and Hermans, J. (1996) *Biopolymers 39*, 479–489.
8. Wimley, W. C., Gawrisch, K., Creamer, T. P., and White, S. H. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 2985–2990.
9. Serrano, L., Horovitz, A., Avron, B., Bycroft, M., and Fersht, A. R. (1990) *Biochemistry 29*, 9343–9352.
10. Horovitz, A., Serrano, L., Avron, B., Bycroft, M., and Fersht, A. R. (1990) *J. Mol. Biol. 216*, 1031–1044.
11. Šali, D., Bycroft, M., and Fersht, A. R. (1991) *J. Mol. Biol. 220*, 779–788.
12. Xiao, L., and Honig, B. (1999) *J. Mol. Biol. 289*, 1435–1444.
13. Anderson, D. E., Becktel, W. J., and Dahlquist, F. W. (1990) *Biochemistry 29*, 2403–2408.
14. Schreiber, G., Buckle, A. M., and Fersht, A. R. (1994) *Structure 2*, 945–951.
15. Honig, B., and Nicholls, A. (1995) *Science 268*, 1144–1149.
16. Davis, M. E., and McCammon, J. A. (1990) *Chem. Rev. 90*, 509–521.
17. Kalia, Y. N., Brocklehurst, S. M., Hipps, D. S., Appella, E., Sakaguchi, K., and Perham, R. N. (1993) *J. Mol. Biol. 230*, 323–341.
18. Spector, S., Kuhlman, B., Fairman, R., Wong, E., Boice, J. A., and Raleigh, D. P. (1998) *J. Mol. Biol. 276*, 479–489.
19. Spector, S., Young, P., and Raleigh, D. P. (1999) *Biochemistry 38*, 4128–4136.
20. Kelley, L. A., Gardner, S. P., and Sutcliffe, M. J. (1996) *Protein Eng. 9*, 1063–1065.
21. Brünger, A. T., and Karplus, M. (1988) *Proteins: Struct., Funct., Genet. 4*, 148–156.
22. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) *J. Comput. Chem. 4*, 187–217.
23. Gilson, M. K., and Honig, B. (1988) *Proteins: Struct., Funct., Genet. 4*, 7–18.
24. Gilson, M. K., Sharp, K. A., and Honig, B. (1988) *J. Comput. Chem. 9*, 327–335.
25. Sharp, K. A., and Honig, B. (1990) *Annu. Rev. Biophys. Biophys. Chem. 19*, 301–332.
26. Hendsch, Z. S., Sindelar, C. V., and Tidor, B. (1998) *J. Phys. Chem. B 102*, 4404–4410.
27. Sitkoff, D., Sharp, K. A., and Honig, B. (1994) *J. Phys. Chem. 98*, 1978–1988.
28. Laue, T. M., Shah, B. D., Ridgeway, T. M., and Pelletier, S. L. (1992) in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S. E., Rowe, A. J., and Horton, J. C., Eds.) pp 90–125, The Royal Society of Chemistry, Cambridge.
29. Braunschweiler, L., and Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521–528.
30. Davies, D. G., and Bax, A. (1985) *J. Am. Chem. Soc. 107*, 2820–2821.
31. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4554.
32. Macura, S., Hyang, Y., Suter, D., and Ernst, R. R. (1981) *J. Magn. Reson. 53*, 259–281.
33. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley and Sons, New York.
34. Shoemaker, D. P., Garland, C. W., and Nibler, J. W. (1989) *Experiments in Physical Chemistry*, 5th ed., McGraw-Hill Publishing Company, New York.
35. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) *J. Biomolecular NMR 5*, 67–81.
36. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) *Protein Sci. 4*, 2138–2148.
37. Pace, C. N. (1992) *J. Mol. Biol. 226*, 29–35.
38. Hendsch, Z. S., Jonsson, T., Sauer, R. T., and Tidor, B. (1996) *Biochemistry 35*, 7621–7625.
39. Pakula, A. A., and Sauer, R. T. (1990) *Nature 344*, 363–364.
40. Spector, S., Rosconi, M., and Raleigh, D. P. (1999) *Biopolymers 49*, 29–40.
41. Mande, S. S., Sarfaty, S., Allen, M. D., Perham, R. N., and Hol, W. G. J. (1996) *Structure 4*, 277–286.
42. Grimsley, G. R., Shaw, K. L., Fee, L. R., Alston, R. W., Huyghues-Despointes, B. M. P., Thurlkill, R. L., Scholtz, J. M., and Pace, C. N. (1999) *Protein Sci. 8*, 1843–1849.
43. Loladze, V. V., Ibarra-Molero, B., Sanchez-Ruiz, J. M., and Makhatadze, G. I. (1999) *Biochemistry 38*, 16419–16423.
44. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946–950.